# THEORETICAL STUDIES OF THE ELECTRONIC STRUCTURES OF GUANINE, CYTOSINE, AND SELECTED DERIVATIVES

A DISSERTATION
SUBMITTED TO THE FACULTY OF ATLANTA UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY

BY
PHELESIA YVONNE JONES-COOPER

DEPARTMENT OF CHEMISTRY

ATLANTA, GEORGIA

JULY 1989

ABSTRACT

CHEMISTRY

COOPER-JONES, PHELESIA Y. B.S. BISHOP COLLEGE 1979
M.S. TEXAS SOUTHERN UNIVERSITY 1982

THEORETICAL STUDIES OF THE ELECTRONIC STRUCTURES OF GUANINE, CYTOSINE, AND SELECTED DERIVATIVES

Advisors: Professors John H. Hall, Jr. and Joe Johnson

Thesis dated July, 1989

*Ab-initio* self-consistent-field (SCF) electronic energies and localized molecular orbitals (LMO's) for the protonated forms of guanine, $m^7$-guanine, the six possible tautomers of guanine, substituted nucleic acid bases and their tautomers are reported. Localized molecular orbital calculations (LMO) are employed to predict the bonding patterns in these molecules.

The electronic energies are also calculated for the five possible tautomers of cytosine and the eighteen possible guanine-cytosine tautomeric base pairs.

Our calculations indicate that the most stable tautomeric form of guanine has the hydrogens at the $N_9$ and $N_1$ positions, and the least stable tautomer is the enol form. The most favorable site of protonation for guanine is predicted to be the $N_7$, and the least favorable site is the $N_3$. The most favorable protonation site for the $m^7$-guanine is predicted to be the $N_3$ site, and the least favorable site is the $O_6$ (trans) site using Partial Retention of Diatomic Differential Overlap (PRDDO) method and ST0-3G basis set.

The calculations for the substituted nucleic acid bases suggest that the presence of a substituent alters tautomerization in cytosine and that hydrogen bonding interactions may be altered via the nearest neighbor effect. Our calculations indicate that the most stable tautomeric form of cytosine has the hydrogen at the $N_1$ position, and that the least stable tautomer is the imino (trans) form.

Our results also indicate that the substituent can affect the relative stability of the hydrogen bonded base pairs. A reaction mechanism is proposed to explain the biological behavior and mutagenicty of $N^4$-substituted cytosine. The most stable guanine-cytosine tautomeric base pair using STO-3G basis set is predicted to be the C3H···G3H9H. These results may be used as a starting point for constructing a theoretical model for macromolecule conformational changes in DNA and RNA polymers which have Watson-Crick and Hoogsteen type hydrogen bonded base pairs.

ACKNOWLEDGEMENTS

The author wishes to express her deepest appreciation to her mentor, Dr. John H. Hall, Jr., for his guidance, patience, understanding, faith, encouragement, and support. She wishes to acknowledge and thank Dr. Joe Johnson for his listening ear, guidance and encouragement. Sincere thanks to Dr. Subhash C. Bhatia for his undying encouragement, tolerance, and support.

The author wishes to thank her parents Dr. and Mrs. Jesse Jones for their believeing in her and who stood by her through the years of her study. The author would like to thank her husband Emanuiel, and her son Mott for their support and allowing her to pursue her dream. The author would also like to thank Gerald Walker for his contribution and his helpful disscussion in preparation of this dissertation. The author would like to thank Andi McDaniel, Vernon Morris, Gerald Walker, Mosumola George-Taylor, and the Alick family for their friendship, which helped her through many difficult times. But, most of all the author would like to give thanks to God for with out whom all of this would not be possible.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

# INTRODUCTION

The deoxyribonucleic (DNA) and ribonucleic (RNA) acids are biopolymers consisting of four major nucleoside monomers: adenosine, guanosine, cytosine and uridine for RNA; and deoxyadenosine, deoxyguanosine, deoxycytosine, and deoxythymidine for DNA. DNA, the hereditary material, is a polymer where monomeric units are linked by phosphodiester bonds at the 3′ and 5′ positions of the ribose moiety. Even if these bases were not constituents of DNA, they would nevertheless be among the most important molecules in biology.[1] For example, alkylation of the N-7 atom of guanine introduces some special chemical properties to the capped 5′ termini of mRNA, but the precise function of this specie is not definite.[2] Watson and Crick first postulated that DNA exists in a double helix form where adenine is base paired to thymine (A-T), and guanine is base paired to cytosine through hydrogen bonding.[3] When the chemical nature of the base is changed by protonation, alkylation, or tautomerization, the hydrogen bonding is altered. The hydrogen bonding in turn alters the base pairing interaction between helical strands. The change in base pairing and subsequent conformational changes can lead to mutations and may stimulate the genesis of cancerous cells or it may also initiate other types of biological response in DNA.[4,5]

There are two classes of mutations: direct and indirect. In the former, spontaneous or induced mispairing,

results in specific base pair substitutions. The indirect mechanism generates mutations in one or more steps which are different from the initial event.[6] There are many different pathways for direct mutation: transitions, transversions, frame-shifts, and deletions.[6] Other investigators have utilized substituted nucleic acid bases[7], interaction of protein with base pairs[8], protonation,[9,10] and tautomerization[11] to probe the mechanism of mutation.

We have employed *ab-initio* self-consistent-field (SCF) methods[12] to investigate the relative energies and bonding patterns of the tautomers of nucleic acid bases and their base pairs because of the low concentration of these tautomers in solution which makes it difficult to study these systems using experimental techniques.[13]

The SCF calculations are performed using Hartree-Fock approximation.[12] This approximation is an attempt to solve the electronic Schrodinger equation for multielectron systems. The simplest antisymmetric wavefunction, which can be used in SCF method to describe the ground state of an N-electron system, is a single determinant wavefunction,

$$| \Psi_o > = | \chi_1 \chi_2 \cdots \chi_n >. \qquad (1)$$

The variation principle states that the best wavefunction of this functional form is the one which gives the lowest

$$E_o = < \Psi_o \mid H \mid \Psi_o > \qquad (2)$$

energy where H is the electronic Hamiltonian. By minimizing $E_o$ with respect to the choice of spin orbitals, one can derive the Hartee-Fock equation, which determines the optimal spin orbitals. The Hartee-Fock equation is nonlinear and must be solved iteratively. The procedure for solving Hartee-Fock equation is called the SCF method. SCF can be achieved by making an initial guess at the spin orbitals, then calculate the average field seen by each electron, and then solve the eigenvalue equations to obtain a new set of spin orbitals. Using these new spin orbitals one can obtain new fields and repeat the procedure until self-consistency is reached.

The electronic wavefunction of a molecule can be used to describe the molecular orbital. The electronic wavefunction depends on the internuclear distance, but the nuclei are assumed to be fixed (Born-Oppenheimer approximation) in determining the electronic wavefunctions. Each molecular orbital can be represented as a linear combination of atomic orbitals (LCAO). Molecular orbital theory assigns electrons to orbitals $\Psi_\iota$ which are linear combinations of a set of basis functions $\phi_\mu$ (eq 3). The $\phi_\mu$ are centered on the atoms

$$\Psi_\iota = \sum_\mu c_{\mu\iota}\phi_\mu \qquad (3)$$

thus expansion of equation (3) is often described as the LCAO approximation.[14] More accurate molecular orbitals can be obtained from larger basis sets which permit increased flexibility in the representation.[14] The most difficult part of LCAO-SCF theory is usually the evaluation of the large number of two-electron integrals.[14] Semiempirical methods usually treat these integrals by approximating their values, thus reducing the computation time. One the other hand, *ab-initio* methods evaluate the integrals but have to use basis functions $\phi_\mu$ for which such integration is possible.[14] To simplify molecular integral evaluation, Boys[15-18] proposed in 1950 the use of Gaussian-type orbitals (GTO's) instead of Slater-type orbitals (STO's) for the atomic orbitals in an LCAO wavefunction. Gaussian integral evaluation requires less computer time than Slater integral evaluation. To get a accurate representation of atomic orbitals one must use a linear combination of several Gaussians.

One important point to note for the interpretation of molecular orbitals is that if p atomic orbitals are used to construct bonding molecular orbitals, the coefficients are of the same sign for a $\pi$ orbital, but the opposite sign for a $\sigma$ orbital.

To determine the valence structures of large molecular systems the localization method introduced by Boys[15-18] is more applicable. The Boys method requires only the one-

electron dipole moment integrals over the occupied molecular orbitals. The Boys localization procedure chooses a transformation T that minimizes the orbitals self-repulsion[19]

$$I(\phi) = \sum_{\iota=1}^{n} \int \phi_\iota(1)\phi_\iota(1) r_{12}^{2} \phi\iota(2)\phi\iota(2) d\upsilon_1 d\upsilon_2 . \qquad (4)$$

Equivalent formulation maximize the sum of squares of the distances between orbitals centroids[19] and maximize the sum of squares of distances of the orbital centroids from the arbitrarily defined origin of the molecular coordinate systems,[19]

$$D(\phi) = \sum_{\iota=1}^{n} [\langle \phi_\iota | \bar{r} | \phi_\iota \rangle]^2 . \qquad (5)$$

Maximization of **D** is easy to implement and requires manipulation of far fewer molecular integrals than would direct minimization of I.

The canonical molecular orbitals (CMO's) are obtained by solving the self-consistent field equations in the approximation of partial retention of diatomic differential overlap (PRDDO).[20] The LMO's are generated by subjecting the CMO's to unitary transformation which maximizes the sum of squares of distance between to centroids of molecular orbitals. The resultant localized molecular orbitals usually correspond to chemical concepts such as lone pairs,

two-center bonds, and inner shells, therefore they are more readily interpreted.[21] The LMO's are used to predict bonding patterns. If the molecular orbital is occupied by N electrons (most often N = 2), the electron population is define as:

$$N\phi^2 = Na^2\chi_a^2 + 2NabS(\chi_a\chi_b/S) + Nb^2\chi_b^2 \qquad (6)$$

where S is the overlap integral between 1s atomic orbitals on the two nuclei a and b, $\phi = a\chi_a + b\chi_b$, $\chi$ is the modified atomic orbitals of the two atoms ($\chi_a$ and $\chi_b$). Upon intergrating over all space, one obtains

$$N = Na^2 + 2NabS + Nb^2 \qquad (7)$$

where the middle term is the overlap population.[22] From the degree of overlap population one can determine the type of bonding. For this study the following criteria are determine by Kleier et al.[21], who observed the following conventions in drawing the LMO's for a tripeptide: (1) for each LMO center (i.e. atom) only centers having population of at least 0.35 e are included in the pictorial representation, (2) centers with electron population density in the range of 0.53 to 0.60 e are conected with dashed lines and (3) all other centers in a given LMO are connected with a solid line. Walker et al. [9] suggest that the

criterion prescribed by Kleier et al.[20] may not be suitable for charged heterocycles with highly delocalizable electron systems.

In this study ab-initio self consistent field calculations[12] were performed to study the charge distribution, relative electronic energies of the tautomeric forms of guanine, cytosine, and guanine-cytosine hydrogen bonded base pairs. The protonation properties of guanine and $m^7$-guanine were also examined. Localized molecular orbital[15-18] (LMO) calculations were performed to predict bonding patterns for the protonated forms of $m^7$-guanine, tautomeric forms of guanine, and the substituted base pairs.

By understanding the monomer-monomer interaction one may be able to build a working model of DNA which would contribute to the understanding of the mechanisms and role of DNA and RNA in cellular control. This dissertation is intended as a contribution to the information needed to make such an evaluation.

## REFERENCES


1. Sanger, W. "Principles of Nucleic Acid Structure"; Springer-Verlag: New York, 1984; Chapter 1.

2. Banerjee, A.K.; Microbiological Reviews, **1980**, 44, 175.

3. Watson, J.D. and Crick, F.H.; Nature, **1953**, 171, 737-738.

4. Morrison, R.T. and Boyd, R.N. "Organic Chemistry 5th edition"; Allyn and Bacon, Inc.: Boston, 1987; Chapter 41.

5. Barton, J.K.; Chemical and Engineering News, **1988**, 66, 30-42.

6. Drake, J.W. and Baltz, R.H.; Biochem. Biophys. Res. Commun., **1976**, 907, 11-35.

7. Negishi, K., Takahash, M., Yamashita, Y., Nishizawa, N., and Hayatsi, H.; Biochemistry, **1985**, 24, 7273.

8. Sarai, A. and Sarito, M.; Int. J. Quantum Chem., **1985**, 28, 399-409.

9. Walker, G.A., Bhatia, S.C. and Hall, J.H.,Jr.; J. Am. Chem. Soc., **1987**, 109, 7629.

10. Del Bene, J.E.; J. Mol. Struct. (Theochem), **1985**, 124, 201-212.

11. Stolarski, R., Kierdaszuk, B., Hagberg, C.E., and Shugar, D.; Biochemistry, **1987**, 26, 4332-4337.

12. Szabo, A. and Ostlund, N.S. "Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory"; Macmillan Publishing Co. Inc.,: New York, 1982; Chapter 3.

13. Wong, P.Y., Wong, K.L. and Kearns, D.R.; Biochem. Biophys. Res. Commum., **1972**, 49, 1980.

14. Pople, J.A.; Molecular Orbital Methods, **1970**, 3, 217.

15. Roothaan, C.C.J.; Rev. Mod. Phys., **1951**, 23, 69.

16. Boyd, S.F.; Rev. Mod. Phys., **1960**, 32, 306.

17. Kleier, D.A., Halgren, T.A., Hall, Jr., J.H. and Lipscomb, W.N.; J. Chem. Phys. **1976**, 61, 3905.

18. Boys, S.F. "Quantum Theory of Atoms, Molecular, and the Solid State"; Academic Press: New York, 1966.

19. Goddard,W.A.III, Dunning, T.H., Jr., and Hunt, W.J.; Chem Phys. Lett., **1970**, 4, 231.

20. Halgren, T.A., and Lipscomb, W.N.; J. Chem. Phys., **1973**,58, 1569.

21. Kleier, D.A. and Lipscomb, W.N.; International Journal of Quantum Chemistry, **1977**, S4, 73.

22. Mulliken, R.S., Ann. Rev. Phys. Chem., **1978**, 29, 12.

# Chapter 1

## A Self-Consistent-Field Study for the Tautomeric and Protonated Forms of Guanine and 7-Methyl Guanine

# INTRODUCTION

The process of protonation and deprotonation at various electronegative sites of nucleic acid bases plays an important role in certain biochemical reactions.[1] This phenomenon often leads to tautomerization of naturally occurring nucleic acid bases. The keto-enol tautomers of guanine are believed to actively participate in the binding mechanism of mRNA to the ribosome which initiates translation. The rate of this translation phenomenon has been shown to be affected by pH.[2] Consequently, changes in pH may lead to stabilization of a given tautomer.

Several tautomeric forms of nucleic acid bases may play important roles in the replication process.[3] The mispairing of nucleic acid bases is believed to be due to the presence of one of the tautomeric forms of a given base. These tautomers are known to induce conformational changes resulting from changes in the type of hydrogen bonding patterns exhibited by these tautomers. Watson and Crick[3] have suggested, as a part of their formulation of replication schemes, that spontaneous mutations might be due to the occasional occurrence of a misplaced purine-pyrimidine pair. In the normal Watson-Crick base pairing scheme, adenine-thymine (A-T) and guanine-cytosine (G-C) base pairs are normally formed. However, the enol tautomer of guanine can pair with uracil, and the imino tautomer of adenine can pair with cytosine.[4] Hoogsteen has described

other possible base pairing schemes.[5]

Base analogs can also substitute for naturally occurring bases in DNA. For example, substituted nucleic acid bases can induce mutations. The 5-bromouracil pairing with guanine causes a transition of A-T to G-C.[6] Negishi et al.[7] also have used $N^4$-aminocytidine to induce A-T to G-C transitions. These authors[7] have proposed two different pathways to account for this transition. In these proposed pathways, the imino tautomer of $N^4$-aminocytosine is paired with adenine, while the amino tautomer of $N^4$-aminocytosine pairs with guanine. Single base pair mutations are commonly used to probe structure and function relationship in DNA and RNA molecules.

Del Bene[8], using the 4-31G basis set, has reported the protonation energies for various protonated forms of guanine. These self-consistent-field molecular orbital (SCF-MO) calculations[8] predict the preferred order of protonation in guanine to be $N_7 > O_6(C_5) > N_1 > N_3 > O_6(N_1)$ (Figure 1-1). Szczepaniak et al.[9] have performed <u>ab-initio</u> molecular orbital calculations on the six possible tautomeric forms of guanine using STO-3G and 3-21G basis sets. The 3-21G basis set calculations predict that the most stable tautomer for the guanine is with hydrogens at the $N_9$ and $N_1$ positions.[9] These latter calculations with the STO-3G basis set predict the enol form to be the most stable (by 8.2 kcal/mole).

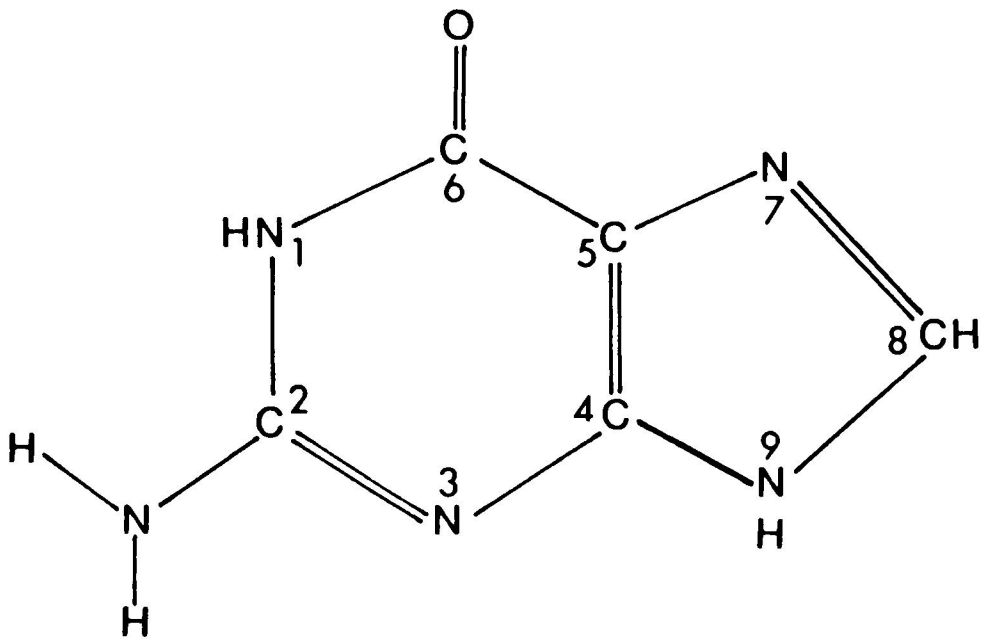



Figure 1-1. Numbering of atoms for guanine and 7-methyl guanine.

To determine the most stable tautomeric form of guanine and 7-methyl guanine ($m^7$-guanine), we have performed SCF calculations using STO-3G and 4-31G basis sets. The relative ease of protonation and deprotonation in guanine and $m^7$-guanine is also investigated by the Partial Retention of Diatomic Differential Overlap (PRDDO) method.[10] The chemical bonding for the tautomers of guanine and $m^7$-guanine is also discussed.

## METHODS

Ab-initio SCF calculations were performed by GAUSS80 program (QCPE No. 446)[11] using STO-3G and 4-31G basis sets.[12-13] The Partial Retention of Diatomic Differential Overlap (PRDDO) method[10] was used to predict the chemical bonding for guanine tautomers and the protonated forms of $m^7$-guanine. The PRDDO calculations employed Slater type orbitals. The localized molecular orbital (LMO) method predicts chemical valence structures by transforming the PRDDO wavefunctions to a localized molecular orbital representation using Boys' criterion.[14-17] The criterion employed to predict the nature of chemical bonds from the LMO's are those described by Kleier et al.[18]

Molecular orbital calculations were performed on the tautomers, the protonated and deprotonated forms of guanine and $m^7$-guanine. The optimized geometries used in these calculations were taken from Del Bene's protonation study of nucleic acid bases.[8] The geometry of the external methyl group, adjacent bond length and bond angles of $m^7$-guanine were optimized by us using the STO-3G basis set (Table 1-1).

TABLE 1-1. The Optimized Geometry for Selected Parameters of 7-Methyl Guanine*

| Atoms | Bond Lengths[a] | Bond Angles[b] | Third Atom for Angle |
|---|---|---|---|
| $H^c$-C | 1.1220 | 109.13 | $N_7$ |
| $H^d$-C | 1.1231 | 110.04 | $N_7$ |
| $H^e$-C | 1.1232 | 110.78 | $H^c$ |
| C-$N_7$ | 1.5248 | 127.56 | $C_5$ |
| $C_5$-$C_4$ | 1.3857 | - | |
| $N_7$-$C_5$ | 1.4563 | - | |
| $C_8$-$N_7$ | 1.3578 | - | |

*The remaining bond lengths and bond angles are from reference 7.
[a]Bond lengths in
[b]Bond angles in degrees
[c]The out-of-plane methyl hydrogen
[d]The in-plane methyl hydrogen
[e]The in-plane methyl hydrogen

## RESULTS and DISCUSSION

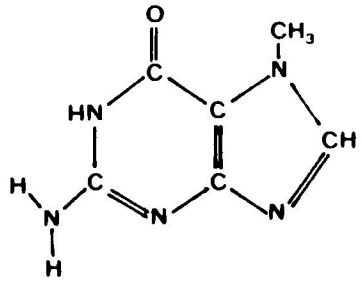
Protonation: The substitution site with greatest electron density is the most probable site for alkylation and protonation in the naturally occurring purine nucleoside. These sites may also represent important sites for enzyme binding.[4,19] Alkylation and protonation of nucleosides are also important in studying phenomena such as hydrogen bonding and base pairing in DNA and RNA.[19] The PRDDO calculations for the protonated species of guanine predict that the $O_6$ site as the most probable site for protonation in the guanine molecule (Table 1-2). However, the energy difference for protonation of the $O_6$ and $N_7$ sites is calculated to be only 7 kcal/mol. Our results also indicate that $N_3$ is the least probable site for protonation (Table 1-2). The PRDDO predicted protonation order ($O_6 > N_1 > N_7 > N_3$) is in agreement with that of Mezey et al.[20] Their minimal basis set (STO-3G) calculations with unoptimized geometries show that the $O_6$ protonation site is preferred by 1.1 kcal/mol over the $N_7$ site. On the other hand, the *ab-initio* SCF calculations, using a more accurate 4-31G basis set, predict that the $N_7$ site is preferred over the $O_6$ site by 11.6 kcal/mol.[8] The energy differences of 11.6 kcal/mol (4-31G) and 7.0 kcal/mol (PRDDO) are both small enough for higher order of calculations to be desirable.

The $m^7$-guanine can have two major tautomeric forms in solution (Figure 1-2; I,II). Addition of a methyl group at

TABLE 1-2. Energy Analysis for Guanine and Protonated Forms of Guanine

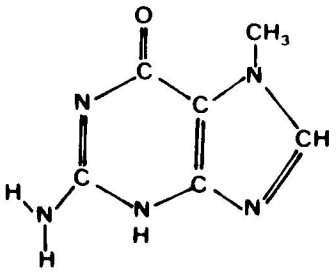
| Molecule | $E^{a}_{PRDDO}$ | ΔE Protonation Energy[b] | | |
|---|---|---|---|---|
| | | PRDDO | PRDDO[c] | 4-31G[d] |
| Guanine (G) | -537.9822 | - | - | - |
| $N_1$-H G[e] | -538.3820 | -250.85 | -275.75 | -220.4 |
| $O_6$-H G | -538.4122 | -269.77 | -282.01 | -233.1 |
| $N_3$-H G | -538.4006 | -262.49 | -266.57 | -228.5 |
| $N_7$-H G | -538.4092 | -267.94 | -275.59 | -244.7 |

[a]The total energy (E) in a.u.
[b]Protonation energy (ΔE) in kcal/mole
[c]Optimized geometry
[d]Protonation Energies from reference 7
[e]$N_1$-H G = site of protonation (ie. at position one of guanine)



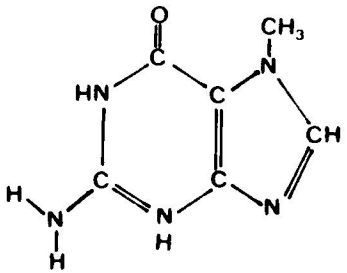
$m^7$-Guanine

I



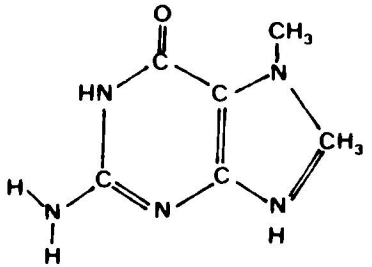
$m^7$-Guanine

II



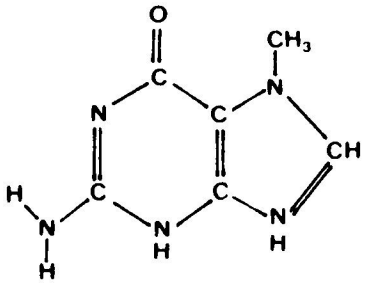
$m^7$-Guanine ($N_3$)

III



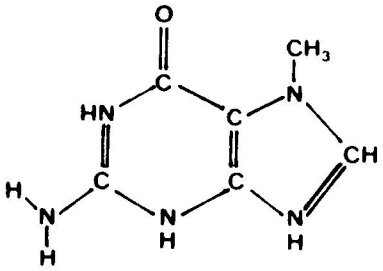
$m^7$-Guanine ($N_9$)

IV



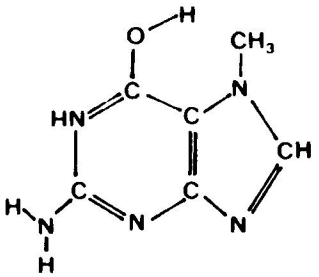
$m^7$-Guanine ($N_3N_9$)

V



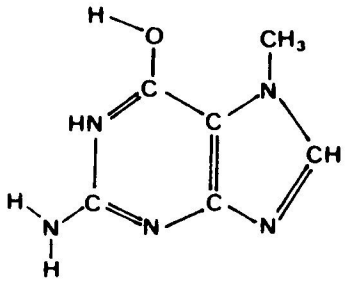
$m^7$-Guanine ($N_3N_9$)*

VI



$m^7$-Guanine (Trans $O_6$)

VII



$m^7$-Guanine (Cis $O_6$)

VIII

*Doubly Protonated



Figure 1-2. Tautomers and protonated forms of 7-methyl guanine

the $N_7$ site of guanine reduces the number of possible sites for protonation from five ($N_7$,$O_6$,$N_3$,$N_9$ and $N_3N_9$) to four ($O_6$,$N_3$,$N_9$ and $N_3N_9$). The energy analysis for $m^7$-guanine is presented in Table 1-3. Both PRDDO and STO-3G level (open shell) calculations predict that the preferred site for protonation is $N_3$. The $M_7N_9$, $M_7N_3$, $M_7O_6$(cis), $M_7O_6$(trans), and $M_7N_3N_9$ have a total of 86 electrons. In the triplet state there are 87 negative eigenvalues, but in the singlet state there are 86 negative eigenvalues; therefore, the lowest lying electronic state is the singlet state. Both levels of approximation predict the energy difference for protonation between the $N_3$ and $O_6$ (trans) site to be 23.5 kcal/mole. The order of protonation to form the doubly protonated form of $m^7$-guanine (Figure 1-2; VI) is via the $N_3$-H protonated form of $m^7$-guanine and is preferred over the $N_9$-H protonated form of $m^7$-guanine by 11.1 kcal/mol (Figure 1-3).

Tautomers: Theoretical calculations have been performed by other investigators[21] to predict which tautomer of guanine is most likely to be involved in spontaneous mutations due to the presence of different tautomeric forms at the time of DNA replication. The six possible tautomeric forms of guanine are shown in Figure 1-4, and the energy analysis is presented in Table 1-4. The PRDDO and the STO-3G minimal basis set calculations predict the relative ordering of the electronic energies for the tautomeric forms of guanine as

TABLE 1-3. Energy Analysis for 7-Methyl Guanine

| Molecules | $E^a$PRDDO | $\Delta E^b$ | $E^a$STO-3G | $\Delta E^b$ |
|---|---|---|---|---|
| $m^7$-Guanine ($m^7$-G) | -576.9040 | 0 | -570.9725 | 0 |
| $N_3$-H $m^7$-G (Tautomer) | -576.8672 | 23.02[c] | -570.9383 | 21.46[c] |
| **Cis** $O_6$-H $m^7$-G | -577.3607 | 115.38 | -571.3701 | 129.37 |
| **Trans** $O_6$-H $m^7$-G | -577.3518 | 109.73 | -571.3840 | 126.29 |
| $N_3$-H $m^7$-G | -577.3960 | 137.46 | -571.3924 | 149.76 |
| $N_9$-H $m^7$-G | -577.3729 | 122.97 | -571.3772 | 138.65 |
| $N_3N_9$-H $m^7$-G | - | - | -571.2794 | 192.57 |

[a]The total energy (E) in a.u.
[b](ΔE) in kcal/mole
[c]Relative tautomeric energy (ΔE)

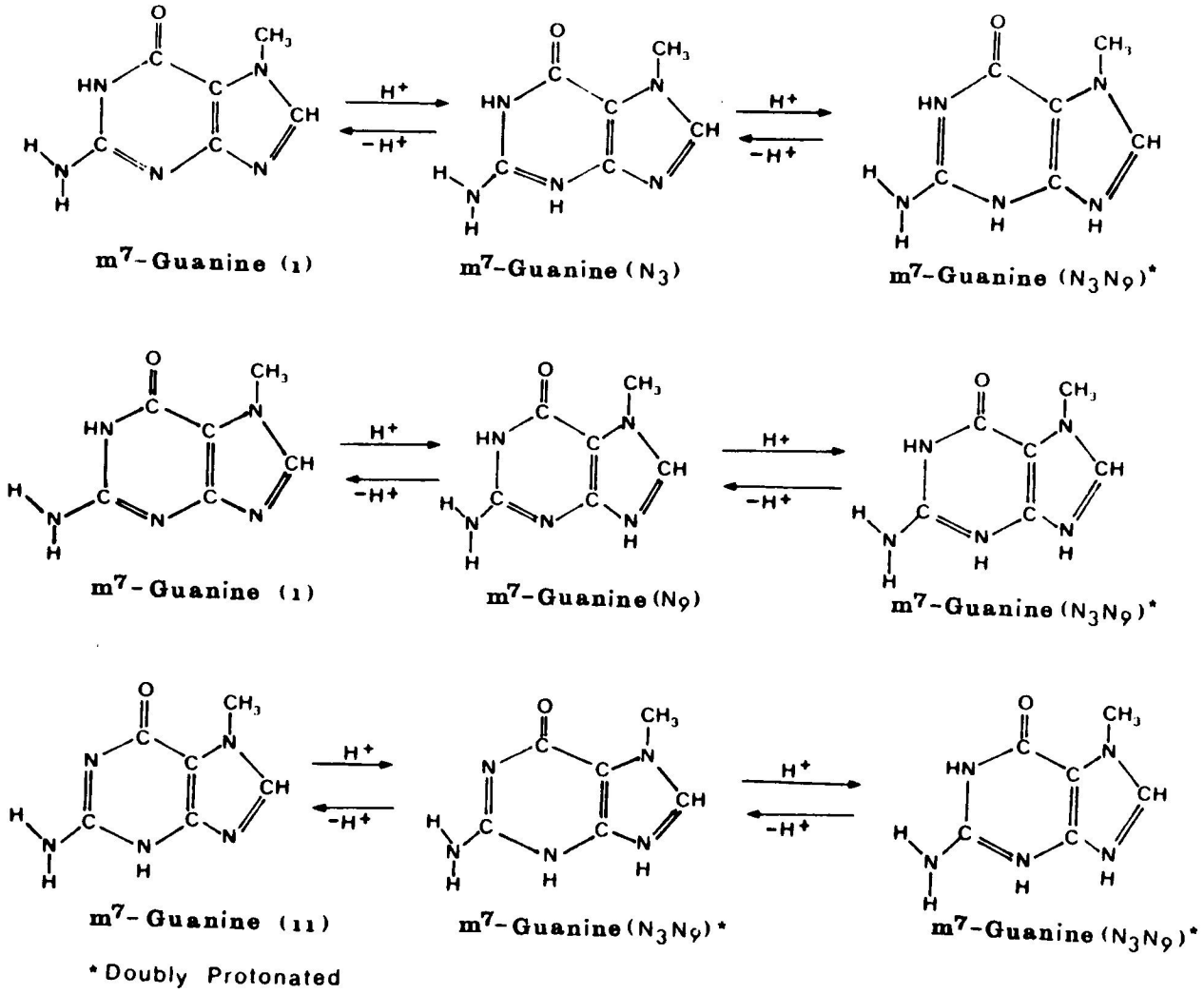



Figure 1-3. Order of protonation for doubly protonated 7-methyl guanine.

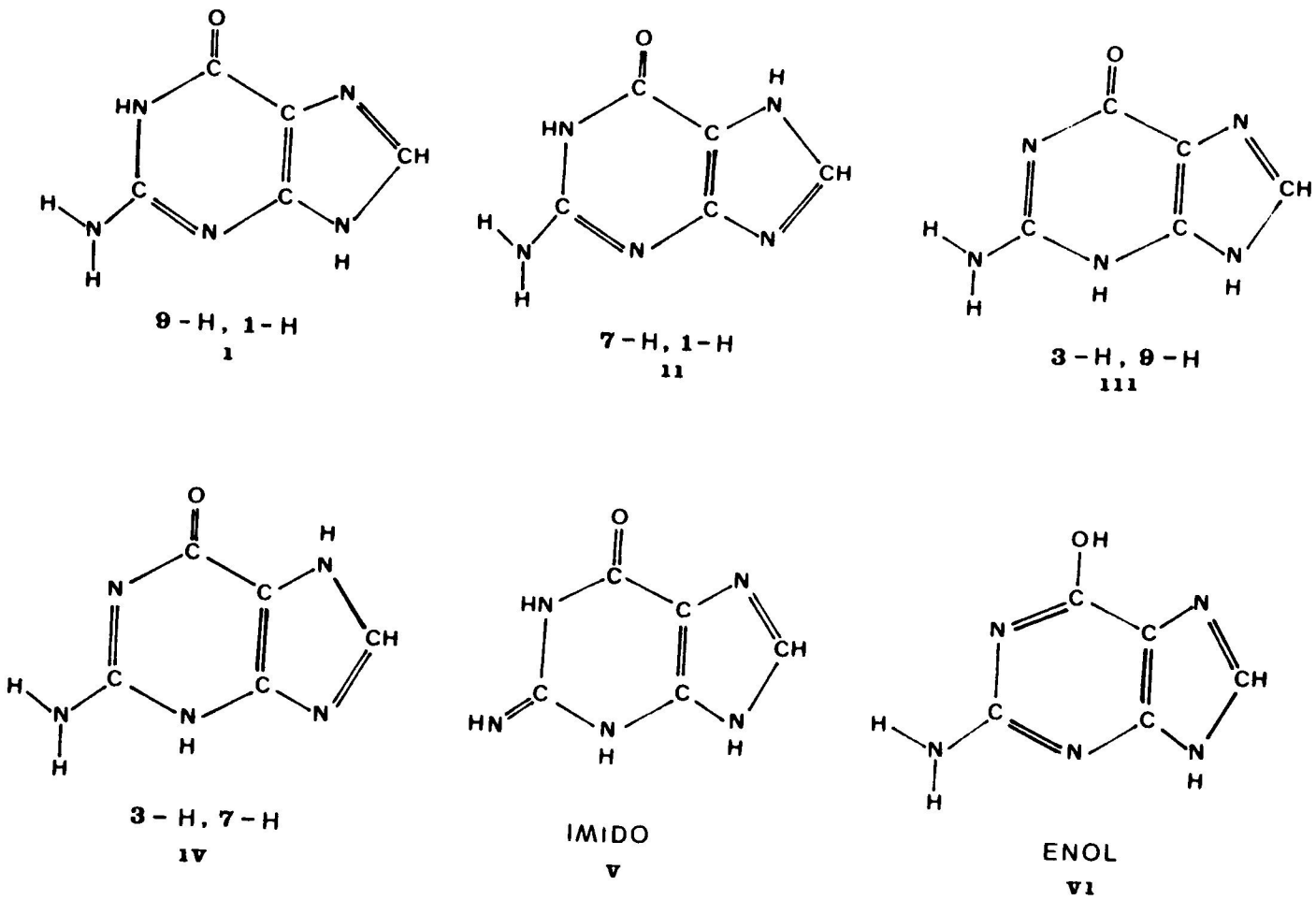





Figure 1-4. Possible tautomers of guanine.

TABLE 1-4. Energy Analysis For The Tautomeric Forms of Guanine

| Tautomeric Forms | Basis Sets | | | | | |
|---|---|---|---|---|---|---|
| | $E^a$ (PRDDO) | $\Delta E^b$ | $E^a$ (STO-3G) | $\Delta E^b$ | $E^a$ (4-31G) | $\Delta E^b$ |
| **9H,1H** | -537.9753 | 0 | -532.4613 | 0 | -538.5567 | 0 |
| **7H,1H** | -537.9529 | 14.49 | -532.4309 | 19.08 | -538.5392 | 10.99 |
| **3H,9H** | -537.9142 | 38.30 | -532.3941 | 42.17 | -538.5023 | 34.10 |
| **3H,7H** | -537.9084 | 41.97 | -532.3828 | 49.29 | -538.5067 | 31.36 |
| **IMINO** | -537.9120 | 39.72 | -532.3906 | 44.39 | -538.5112 | 28.53 |
| **ENOL** | -537.9300 | 28.42 | -532.4092 | 32.72 | -538.4969 | 37.53 |

[a]The total energies (E) in a.u.
[b]The relative energy ($\Delta E$) in kcal/mole

9H,1H < 7H,1H < enol < 3H,9H < imino < 3H,7H. It is not surprising that relative energies calculated using PRDDO and STO-3G methods are similar since these calculations are at the same level of approximation.

Szczepaniak et al.[9] employing a 3-21G basis set predicted the relative ordering for the electronic energies of the three tautomers to be 9H,1H < 7H,1H < enol. Our calculations, with the 4-31G basis set, predict the relative ordering of the energies as 9H,1H < 7H,1H < imido < 3H,7H < 3H,9H < enol (Table 1-4). Latajka et al.[22] using the 3-21G basis set also predict the 9H,1H tautomer of guanine to be the most stable form. The 7H,1H tautomer of guanine is only 2.6 kcal/mol higher in energy, and the enol tautomer of guanine is predicted to be higher in energy by 5.0 kcal/mol.[22] The X-ray crystallographic studies of guanine monohydrate indicate that in hydrated crystals, the 9H,1H tautomer of guanine is the most stable form.[23] Our calculations suggest that the 9H,1H tautomeric form of guanine is the most stable and should be the predominant form in transitions involving guanine as the initial nucleoside.

7-Methyl guanosine has been shown to exist in a zwitterionic form, a keto form, and an enolate form.[24] Near neutral pH, the two forms exist in almost equal quantities ($p_{ka}$ = 6.7 - 7.2 for $m^7$-GMP). Laser Raman studies performed by Pambi[24] indicate the presence of the keto and enolate

form of $m^7$-GMP. These measurements are in agreement with the experimental observation that an increase in hydrogen (or deuterium) ion concentration will shift the equilibrium towards the keto form of guanine.[24] It has been suggested[2] that the enolate form of 7-methyl guanosine exists in the cap structure which preferably complexes with the cap-binding protein. The PRDDO and the STO-3G calculations predict the enolate form of $m^7$-guanine to be less stable than the keto form of $m^7$-guanine by 22.24 kcal/mol. Thus, at this level of approximation, it is predicted that the keto form of $m^7$-guanine should complex with the cap-binding protein for the initiation of protein synthesis. The Raman studies in $D_2O$ solution do not indicate significant protonation of $O_6$ occurring simultaneously with the equilibrium induced protonation of $N_1$.[24] Thus, solvent effects may play a major part in stablizing a given tautomer.[25]

Deprotonation of guanine occurs over a small range of pH.[24] At a pH value of approximately 10.5, a hydroxyl ion ($OH^-$) is proposed to react with the $m^7$-guanine at the $C_8$ position causing the imidazole ring to open; this in turn causes "denaturation" of $m^7$-GMP.[24] The calculated bond orders from the PRDDO wavefunction indicate that the $C_4$-$N_9$ bond is the weakest bond in the imidazole ring of $m^7$-guanine (Table 1-5), while the weakest bond in the imidazole ring of guanine is the $C_8$-$N_9$ bond (Table 1-6). It appears that the

TABLES 1-5. Armstrong-Perkins-Stewart Population Analysis for 7-Methyl Guanine

| Bonds | Bond Order |
|---|---|
| N-H[a] | 0.964 |
| N-H[b] | 0.958 |
| C-H[c] | 0.981 |
| C-H[d] | 0.973 |
| C-H[e] | 0.982 |
| $C-N_7$ | 0.982 |
| $C_2-N_1$ | 1.149 |
| $C_2-N$ | 1.164 |
| $C_2-N_3$ | 1.518 |
| $C_4-C_5$ | 1.423 |
| $C_4-N_3$ | 1.286 |
| $C_4-N_9$ | 1.118 |
| $C_5-N_7$ | 1.132 |
| $C_6-N_1$ | 1.001 |
| $C_6-C_5$ | 1.049 |
| $C_6-O$ | 1.859 |
| $C_8-N_7$ | 1.180 |
| $C_8-H$ | 0.969 |
| $C_8-N_9$ | 1.698 |
| $N_1-H$ | 0.958 |

[a]The amino hydrogen trans to $C_2-N_3$
[b]The amino hydrogen cis to $C_2-N_3$
[c]The out-of-plane methyl hydrogen
[d]The in-plane methyl hydrogens
[e]The in-plane methyl hydrogens

TABLE 1-6. Armstrong-Perkins-Stewart Population Analysis for Guanine

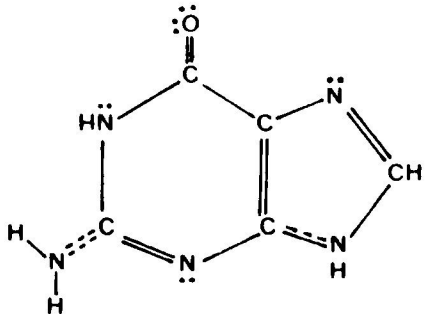
| Bonds | Bond Orders |
|---|---|
| $N-H^a$ | 0.961 |
| $N-H^b$ | 0.957 |
| $C_2-N$ | 1.199 |
| $C_2-N_1$ | 1.191 |
| $C_4-N_9$ | 1.158 |
| $C_5-C_4$ | 1.387 |
| $C_5-N_7$ | 1.182 |
| $C_6-O$ | 1.876 |
| $C_6-C_5$ | 1.040 |
| $C_8-N_7$ | 1.702 |
| $C_8-N_9$ | 1.152 |
| $C_8-H$ | 0.971 |
| $N_1-C_6$ | 0.996 |
| $N_1-H$ | 0.955 |
| $N_3-C_4$ | 1.288 |
| $N_3-C_2$ | 1.452 |
| $N_9-H$ | 0.955 |

[a]The amino hydrogen trans to $C_2-N_3$
[b]The amino hydrogen cis to $C_2-N_3$

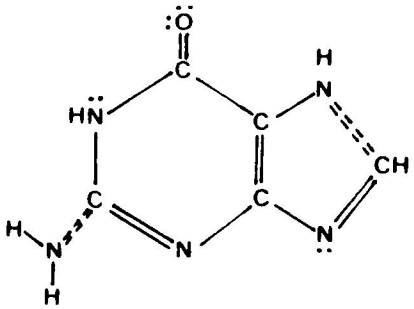
presence of methyl group ($-CH_3$) at the $N_7$ position redistribution of electrons which would otherwise participate in the $\pi$ bonding orbitals of the imidazole ring of $m^7$-guanine. This redistribution of electronic density destabilizes the imidazole ring making the $C_4$-$N_9$ bond more susceptible to nucleophilic attack.

The methyl group at the $N_7$ position of $m^7$-guanine imparts special chemical properties to the capped 5'-termini of mRNA's, but the precise function of the 7-methyl group is not clear.[26] Experimental studies by Jones et al.[19] report that the $m^7$-guanosine readily loses D-ribose under the influence of acid and/or heat to give $m^7$-guanine; whereas, guanosine does not loses its D-ribose group under the same conditions.[19] Our calculated bond order for $m^7$-guanine suggests that the $C_4$-$N_9$ bond (Figure 1-1) should be more susceptible to attack by hydroxyl group and heat. The attachment of D-ribose at $N_9$ position may further weaken this bond and as well as glycosyl bond in $m^7$-guanosine. Therefore, our results are in agreement with the observations of Jones et al.[19]

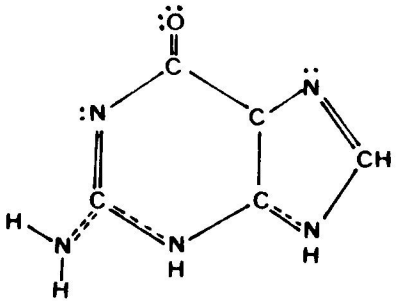
Chemical Bonding: The bonding patterns using localized molecular orbitals for the tautomeric forms of guanine are shown in Figure 1-5. The LMO center having an electron population density (EPD) of < 0.35 electrons (e) is not considered to be bonding.[18] The LMO centers having 0.35 e to 0.60 e are drawn as dotted lines, and those having EPD of



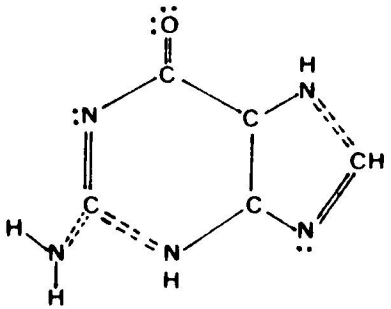
9-H, 1-H
I



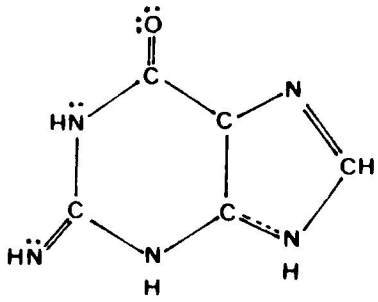
7-H, 1-H
II



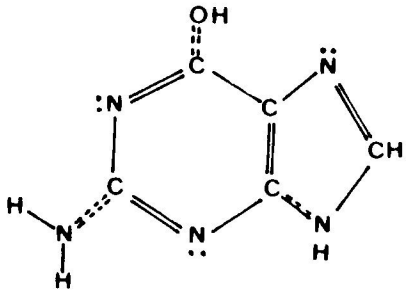
3-H, 9-H
III



3-H, 7-H
IV



IMIDO
V



ENOL
VI

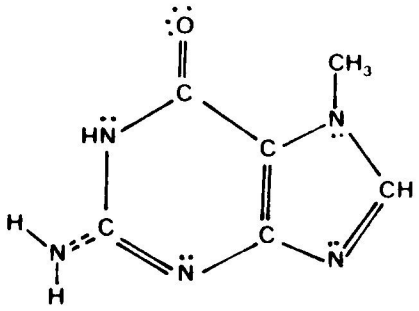


Figure 1-5. Localized molecular orbital structure for guanine tautomers.

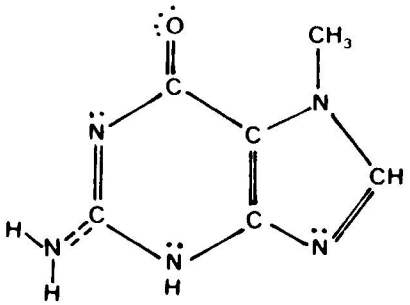
> 0.60 e are drawn as solid lines.[18] Hydrogen atoms attached to nitrogen are able to migrate to other free nitrogens or to the keto oxygen within the same molecule. When the hydrogens migrate to other free heteroatoms within the molecule, there is a shift of electron density as shown by the localized valence structures (Figure 1-5). This redistribution of electron density is observed for all of the localized valence structures of the guanine tautomers. It appears that the redistribution of electrons is dependent upon which heteroatom the hydrogen migrates to and that in turn depends on the available heteroatom which has the highest electron density. The bonding patterns of the guanine tautomers indicate that the nature of chemical bonding is dependent on the location of the hydrogen atoms. A similar behavior has been observed for the tautomeric forms of adenine.[27] A three center bond is predicted for $O_6$-H (**cis**) tautomer to form between $C_4$, $C_5$ and $C_6$ atoms, althought it did not meet the three center bonding criterion.[18] But, since bond for $C_4$ and $C_5$ (population of 1.06 e and 0.61 e ,respectively) did meet the two center bonding criterion, a two center bond is drawn between the $C_4$ and $C_5$ instead of a three center bond between $C_4$, $C_5$ and $C_6$ atoms. On the other hand, Kleier et al.[18] have predicted a three center bond for the guanine molecule between $N_9$,$C_4$ and $C_8$ atoms using PRDDO optimized geometry.

The localized molecular orbitals for the protonated

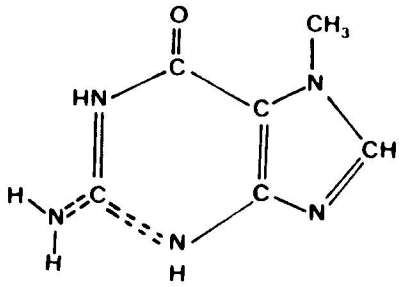
forms of $m^7$-guanine are presented in Figure 1-6. A general trend of all the localized molecular orbital structures (I-VI) is the partial double bond between $C_2$ and the amino nitrogen. The LMO bonding patterns indicate that the nature of chemical bonding is again dependent on the location of the hydrogen atom at the protonation site. The LMO structure of the $N_9$-H tautomer of $m^7$-guanine (Figure 1-6, IV) has five bonds at the $C_8$ position. This phenomenon is also observed in the protonated forms of adenine.[27] The additional electrons at the $C_8$ position are due to charge redistribution. Del Bene[8] has observed that charge transfer of the protonated DNA bases occurs through the sigma electron system and has a stabilizing effect. Dyczmons et al.[28] have calculated the energy and equilibrium geometry of $CH_5^+$ near the Hartree-Fock limit. Their calculations[28] predict that two of the hydrogens participate in a three-atom, two-electron bond with the carbon while the other three hydrogens are normal carbon-hydrogen bonds giving rise to the $C_S$ symmetry. Sefcik et al.[29] employed ion cyclotron resonance spectroscopy to test the chemical equivalency of the methanium ion ($CH_5^+$) protons. Their[29] results suggest that the methanium ion has $C_s$ symmetry as predicted by Dyczmons et al.[28] The predicted five bonds at the $C_8$ position are possible as supported by experimential and theoretical calculations on $CH_5^+$. Another alternative explanation by Walker et al.[27] is that the criterion



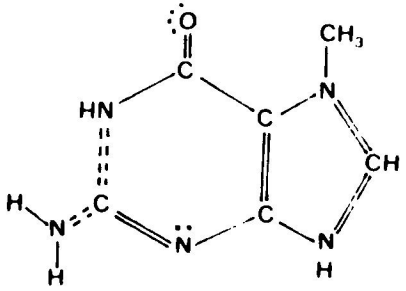
$m^7$-Guanine
I



$m^7$-Guanine
II

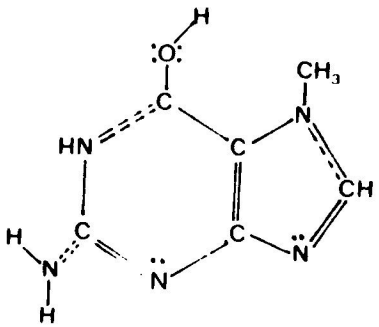


$m^7$-Guanine ($N_3$)
III

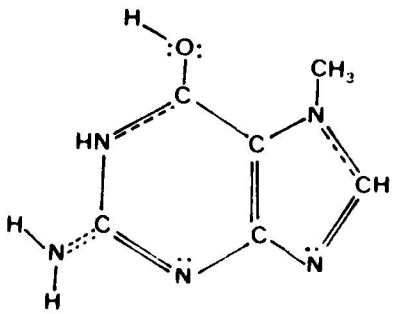


$m^7$-Guanine ($N_9$)

IV



$m^7$-Guanine
(Trans $O_6$)
V



$m^7$-Guanine
(Cis $O_6$)
VI



Figure 1-6. Localized molecular orbital structure for protonated and tautomeric forms of 7-methyl guanine.

suggested for predicting bonding patterns in peptides by Kleier et al.[18] may not accurately predict the bonding pattern for charged heterocycles with highly delocalizable electron systems.

## CONCLUSIONS

Our calculations using PRDDO predict the most favorable site of protonation for guanine to be the $N_7$ and the least favorable site to be $N_3$. Our calculations using PRDDO, STO-3G, and 4-31G basis sets also predict the most stable tautomeric form of guanine to be the 9H,1H and the least stable to be the 3H,7H form. These calculations using PRDDO and STO-3G basis set suggest that the $N_3$ site for $m^7$-guanine is the preferred site. The localized molecular orbital calculations using PRDDO for guanine and $m^7$-guanine indicate that the bonding pattern for their tautomeric forms is dependent on the protonation site. The $C_4N_9$ bond is the weakest bond in $m^7$-guanine and therefore may be responsible for imidazole ring opening in strongly basic solutions.

REFERENCES


1. Vebrich, T.L.Y."Purine, Pyrimidines and Nucleotides and the Chemistry of Nucleic Acids"; Pergamon Press: New York, 1964.

2. Rhoads, R.E., Hellman, G.M., Rremy, P., and Ebel, P.J.; Biochemistry, **1983**, 22, 6084.

3. Watson, J.D., and Crick, H.C.; Nature, **1953**, 171, 964.

4. Saenger, W.; Principles of Nucleic Acid Structure, Springer-Verlag, New York, 1974.

5. Hoogsteen, K.; Acta Crystallogr., **1963**, 16, 907.

6. Styer, L. "Biochemistry"; W.F. Freeman and Company: San Francisco, 1981.

7. Negishi, K., Takahashi, M., Yamashita, Y., Nishizawa, N., and Hayatsi, H.; Biochemistry, **1985**, 24, 7273.

8. Del Bene, J.E.; J. Phys. Chem., **1983**, 87, 367.

9. (a) Szcepaniak, K., Latajka, Z., Morokuma K., and Person, W.B.; Proceedings from the V$^{th}$ International Conference on the Spectroscopy of Matrix Isolated Species, Abbaye de Fontevraud, France, 1985.

   (b) Szczepaniak K., and Szczesniak, M.; J. Mol. Struct. (Theochem), **1987**, 156, 29.

10. Halgren, T.A., and Lipscomb, W.N.; J. Chem. Phys., **1973**,58, 1569.

11. Singh, C., and Kollman, P.; Quantum Chemistry Program Exchange (QCPE), Program No. 446, Department of Pharmaceutical Chemistry, University of California at San Francisco.

12. Hehre, W., Stewart, R., and Pople, J.; J. Chem. Phys., **1969**, 51, 2657.

13. Ditchfield, R., Hehre, W., and Pople, J.; J. Chem. Phys., **1971**, 54, 724.

14. Boys, S.F.; Rev. Mod. Phys., **1968**, 32, 306.

15. Foster, J.M., and Boys, S.F.; Rev. Mod. Phys., **1960**, 32, 300.

16. Boys, S.F. "Quantum Theory of Atoms, Molecular, and the Solid State"; Academic Press: New York, 1966.

17. D.A. Kleier, T. Halgren, J.H. Hall Jr.,and W. Lipscomb, J. Chem. Phys., **1974**, 61, 3905.

18. Kleier, D.A., and Lipscomb, W.N.; presented at the International Journal of Quantum Chemistry: Quantum Biology Symposium, **1977**, 4,73.

19. Jones, J.W., and Robins, R.K.; J. Am. Chem. Soc., **1963**, 59, 193.

20. Mezey,.P.G.; Theoret. Chim. Acta (Berl.), **1980**, 54, 251.

21. Pullman, B., and Pullman, A. " Quantum Biochemistry"; Interscience Publisher: New York, 1963.

22. Latajka, Z., Person, W.B., and Morokuma, K.; J. Mol. Struct. (Theochem), **1986**, 135, 253.

23. Tables of Interatomic Distance and Configuration in Molecules and Ions, The Chemical Society, Burlington House, London, 1965.

24. Pambi, D.S.; M.S. Thesis Atlanta University, 1988 and to be submitted for publication.

25. Sygula, A., and Buda, A.; J. Mol. Struct., **1983**, 92, 267.

26. Banerjee, A.K.; Microbiological Reviews, **1980**, 44, 175.

27. Walker, G.A., Bhatia, S.C., and Hall Jr., J.H.; J. Am. Chem. Soc., **1987**, 109, 7629.

28. Dyczmons, V., Staemmler V., and Kutzelnigg, W.; Chem. Phys. Lett., **1970**, 5, 361.

29. Sefcik, M.D., Henis, J.M.S., and Gaspar, P.P.; J. Chem. Phys., **1974**, 61, 4321.

# Chapter 2

## A study of the Watson-Crick and Non-Watson-Crick Base Pairs*

*See Appendix

# INTRODUCTION

There are many different types of mutations.[1] A transition is a point mutation which occurs by the replacement of a purine with another purine or a pyrimidine with another pyrimidine (Figure 2-1) in a nucleic acid sequence.[1] Transitions are believed to occur due to the presence of tautomers.[2] Substituted nucleosides have been used as probes for studying mutagenic mispairing reaction mechanisms[3] because the substituents on the rings of the nucleic acid bases affect the rate of tautomerization.[4] Experimental studies have shown that 2-aminopurine and 5-bromouracil cause opposing transitions in Φ80,[5] and that $N^4$-aminocytidine induces the adenine:thymine to guanine:cytosine (A:T ---> G:C) transition.[6] $N^4$-hydroxycytidine has been shown to cause G:C ---> A:T transitions[7] in E. Coli, S. typhimurium cells[7] and bacteriophage Qβ,[8] whereas the A:T ---> G:C transition is reported in plasmid PβG.[9] Thus, the transition induced in DNA by a substituted nucleic acid base varies with the parameters of the system. A more thorough understanding of this phenomenon may also lead to an explanation of the observed variable behavior of pharmeceuticals in patients at the molecular level.

The nucleic acid base sequence may serve as a biological identification code for enzymes as well as for the initiation of biological reactions. Nucleic acid base

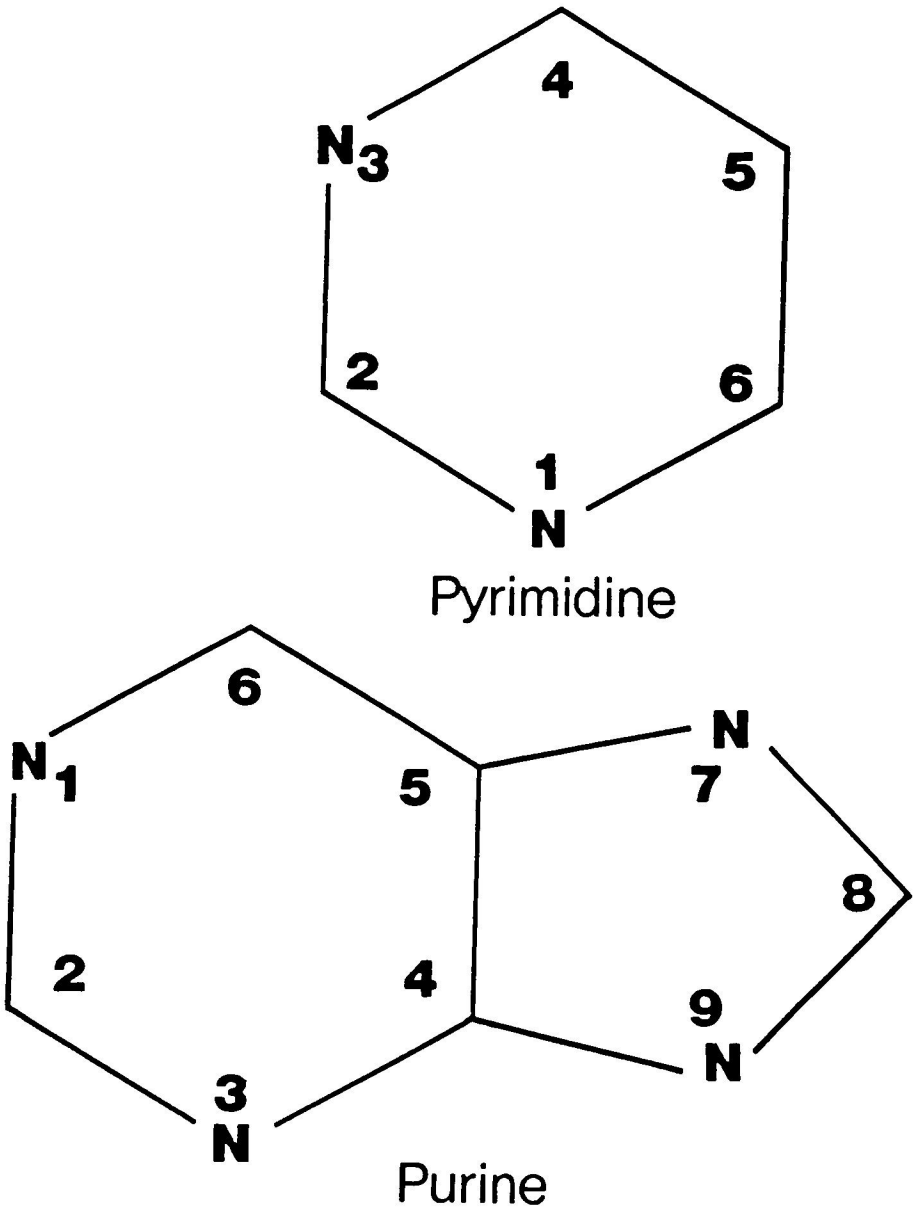



Figure 2-1. Numbering of atoms for purine and pyrimidine ring systems.

sequences also code for the peptide primary sequence protein and a host of other biological (or cellular) responses. Thus, changes in the sequence (eg., due to transitions) may significantly affect the code for a protein, or it may change an important binding site for an enzyme. Changes in the base-base interactions, changes in the electronic distribution in the rings of the base, and changes in substituent-enzyme interactions due to the presence of various substituents on the rings of the bases may also affect cellular behavior.

The low concentration of the mutagenic tautomer in solution makes it difficult to study the mechanisms of these transitions by experimental techniques, so theoretical methods are employed to assist in elucidating these biological reaction mechanisms. Previous work in our laboratory investigated tautomerization in adenine and guanine bases by experimental and theoretical methods.[10] In order to extend these studies, we have performed *ab initio* self-consistent field (SCF) calculations on cytosine (C), $N^4$-substituted cytosines and their tautomers. The relative electronic energy differences between the tautomers, individual bases, H-bonded base pairs, and the localized molecular structures for the tautomers and base pairs are examined to explain the effect of $N^4$-substitution on tautomerization and on the Watson-Crick and non-Watson-Crick H-bonding interactions of C and $N^4$-substituted cytosine ($C^X$)

with guanine (G) and adenine (A). Possible mechanisms for the A:T ---> G:C and G:C ---> A:T transitions are proposed.[6]

## METHODS

*Ab initio* SCF calculations[11] were performed on cytosine (C), $N^4$-hydroxycytosine ($C^{OH}$), $N^4$-aminocytosine ($C^{Am}$), their tautomers, and $N^4$-methylcytosine ($C^{Me}$) (Figure 2-2). The H-bonding interactions for the adenine:$N^4$-substituted-cytosine (A:$C^X$), guanine:$N^4$-substituted-cytosine (G:$C^X$) (Figure 2-3), adenine:thymine (A:T), and adenine:uracil (A:U) base pairs were also investigated. Optimized geometries for all of the bases were obtained from Del Bene's study of base protonation[12] except for the adenine:uracil H-bonding interaction.[13] The optimized geometries for A, U and the A:U pair were obtained from the study of Ohta et al.[13] The GAUSS80 program[14] was used with STO-3G[15] and 4-31G[16] basis sets. The geometries of the substituted bases were optimized in the region of substitution (Figure 2-2). The H-bonding distance and angles (Figure 2-3) were optimized with the STO-3G basis set. The localized molecular orbitals[17] were obtained using the Partial Retention of Diatomic Differential Overlap (PRDDO) method.[18]

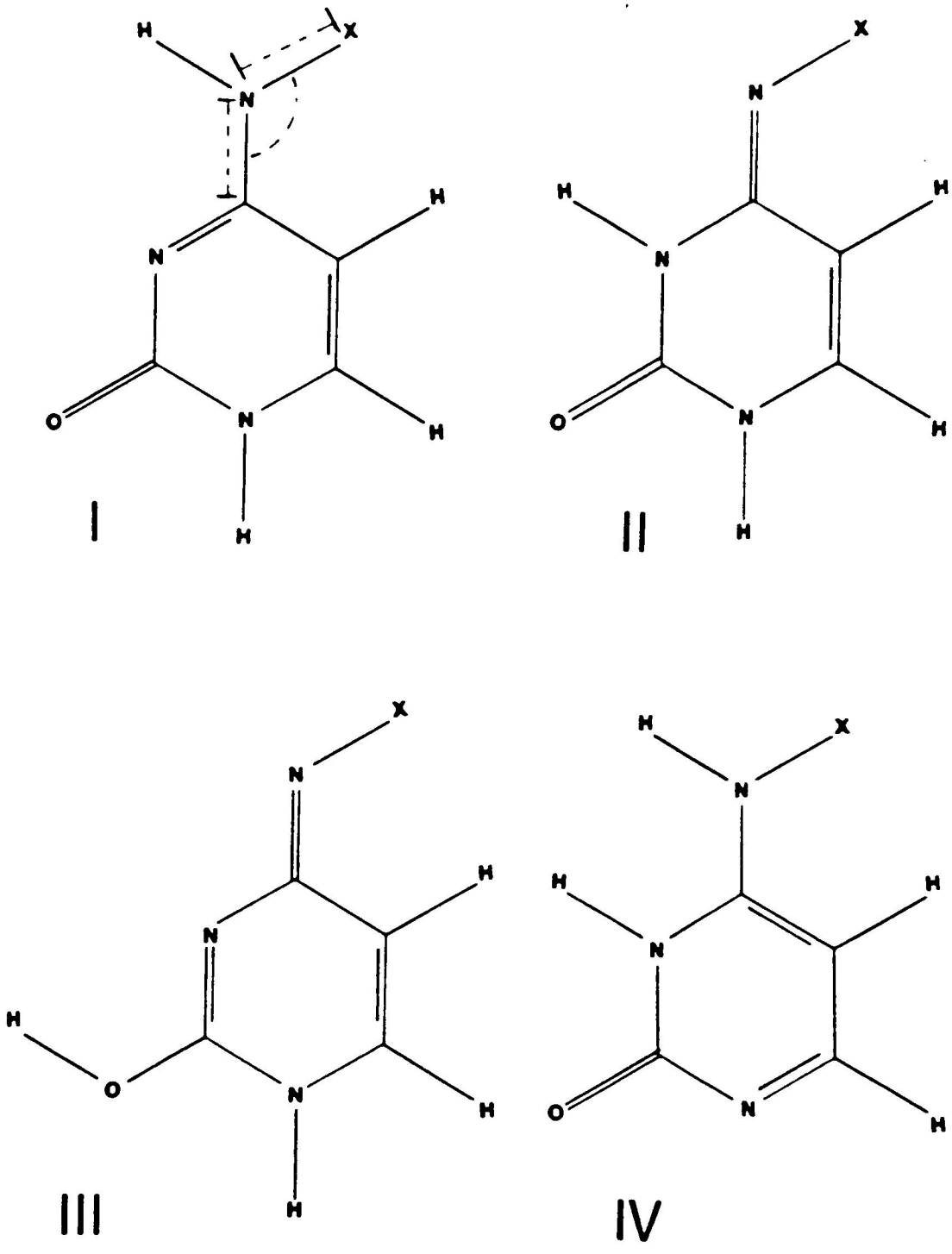



Figure 2-2. Bond lengths and bond angles to be optimized of $N_4$-Cytosine and tautomers.

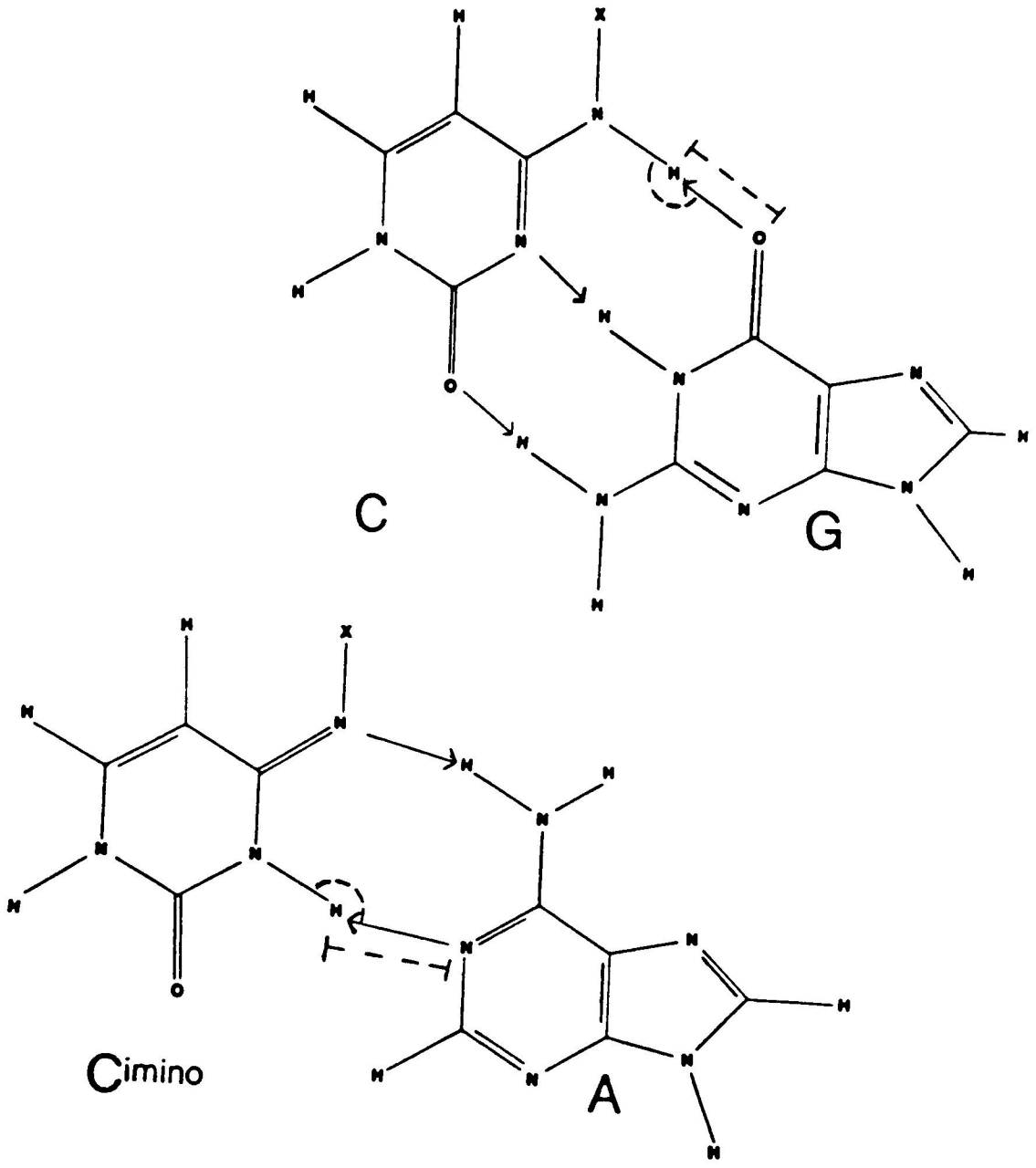



Figure 2-3. Bond lengths and bond angles to be optimized of H-bonding scheme for G:$C^X$ and A:$C^X$.

## RESULTS and DISCUSSION

### Tautomer Stabilities

The electronic energies for the $C^X$ (where X = H, $CH_3$ $NH_2$, and OH) series are listed in Table 2-1. The results show that tautomer I is the most stable tautomer for each species under investigation. The calculated relative stability for $C^{OH}$ at the STO-3G level shows that tautomer III is more stable than tautomer II by 1.07 kcal/mole (Table 2-1). Since our calculated order of tautomer stability for $C^{OH}$ is not consistent with the results obtained for C and $C^{Am}$, we performed calculations on $C^{OH}$ and its tautomers with the 4-31G basis set. Calculations performed with an extended basis set (4-31G) using STO-3G optimized geometries are reported to give more reliable relative energies than those obtained at the minimal basis set level.[19] The results for $C^{OH}$ using the 4-31G basis set indicate that tautomer II is more stable than tautomer III by 5.3 kcal/mole (Table 2-1).

The relative energies of tautomers I and II ($\Delta E_{I-II}$) for $N^4$-substituted cytosines are listed in Table 2-2. The $\Delta E_{I-II}$'s for C, $C^{OH}$, and $C^{Am}$ are 14.6, 11.7 and 2.5 kcal/mole, respectively. Our results show a significant decrease in $\Delta E_{I-II}$ for $C^{Am}$. Therefore, tautomer II of $C^{Am}$ is expected to be present in solution in larger amounts than tautomer II of $C^{OH}$ or C. The lower $\Delta E_{I-II}$ for $C^{Am}$ may be an indication why $C^{Am}$ is more mutagenic than $C^{OH}$.[3]

TABLE 2-1. Electronic Energies[a]

| Tautomer | Basis Set | C | $C^{OH}$ | $C^{NH_2}$ | $C^{CH_3}$ |
|---|---|---|---|---|---|
| I | STO-3G | -387.5453 | -461.3226 | -441.8327 | -426.1048 |
| | PRDDO | -391.5453 | -466.0609 | -446.3888 | - |
| | 4-31G | - | -466.6540 | - | - |
| II | STO-3G | -387.5219 | -461.3038 | -441.8287 | -426.0982 |
| | PRDDO | -391.5284 | -466.0495 | -446.3877 | - |
| | 4-31G | - | -466.6253 | -446.9107 | - |
| III | STO-3G | -387.4980 | -461.3055 | -441.8049 | - |
| | PRDDO | - | - | - | - |
| | 4-31G | - | -466.6168 | - | - |

[a]Energies are in atomic units.
C = Cytosine
$C^{OH}$ = $N^4$-hydroxycytosine
$C^{NH_2}$ = $N^4$-aminocytosine
$C^{CH_3}$ = $N^4$-methylcytosine

TABLE 2-2. Tautomerization Energy[a] ($\Delta E_{I-II}$)

| Tautomer | Basis Set | C | $C^{OH}$ | $C^{CH_3}$ | $C^{NH_2}$ |
|---|---|---|---|---|---|
| I | STO-3G | 0 | 0 | 0 | 0 |
| | PRDDO | | | | |
| | 4-31G | | | | |
| II | STO-3G | 14.6 | 11.7 | 4.1 | 2.5 |
| | PRDDO | 10.6 | 7.2 | - | 0.7 |
| | 4-31G | - | 18.0 | - | - |

[a]Energies are in kcal/mole.
C = Cytosine
$C^{OH}$ = $N^4$-hydroxycytosine
$C^{NH_2}$ = $N^4$-aminocytosine
$C^{CH_3}$ = $N^4$-methylcytosine

A plot of atomic number of the central atom of the substituent versus $\Delta E_{I-II}$ for the cytosine series is shown in Figure2-4. This plot indicates that $\Delta E_{I-II}$ can be lowered or raised by altering the central atom of the substituent. The tightness with which the electrons are held also varies with the substituent and should therefore affect the ease of movement of electrons around the ring. The results of Bernasconi et.al.[20] indicate that synchronized electron motion may play a major role in determining intrinsic barriers.[21] Our results suggest that synchronized electron motion can be altered by the substituent, and that alterations in synchronized electron motion may affect relative tautomer stabilities. Cieplak, et. al.[22] report that the location of the substituent is also important in determining relative tautomer stabilities. Thus, the ability to direct the interaction between nucleic acid bases should be possible by placing the proper substituent in the proper position. One must also consider the environmental effects because the substituents' hyphobicity or hydrophilicity may alter the diffusion and binding properties of the system as well as important enzyme-substrate interactions.

H-bonded Complexes

The H-bonding energies for the base pairs considered are listed in Table 2-3. The tautomers and base pairs under investigation are shown in Figures 2-5,2-6, and 2-7. The

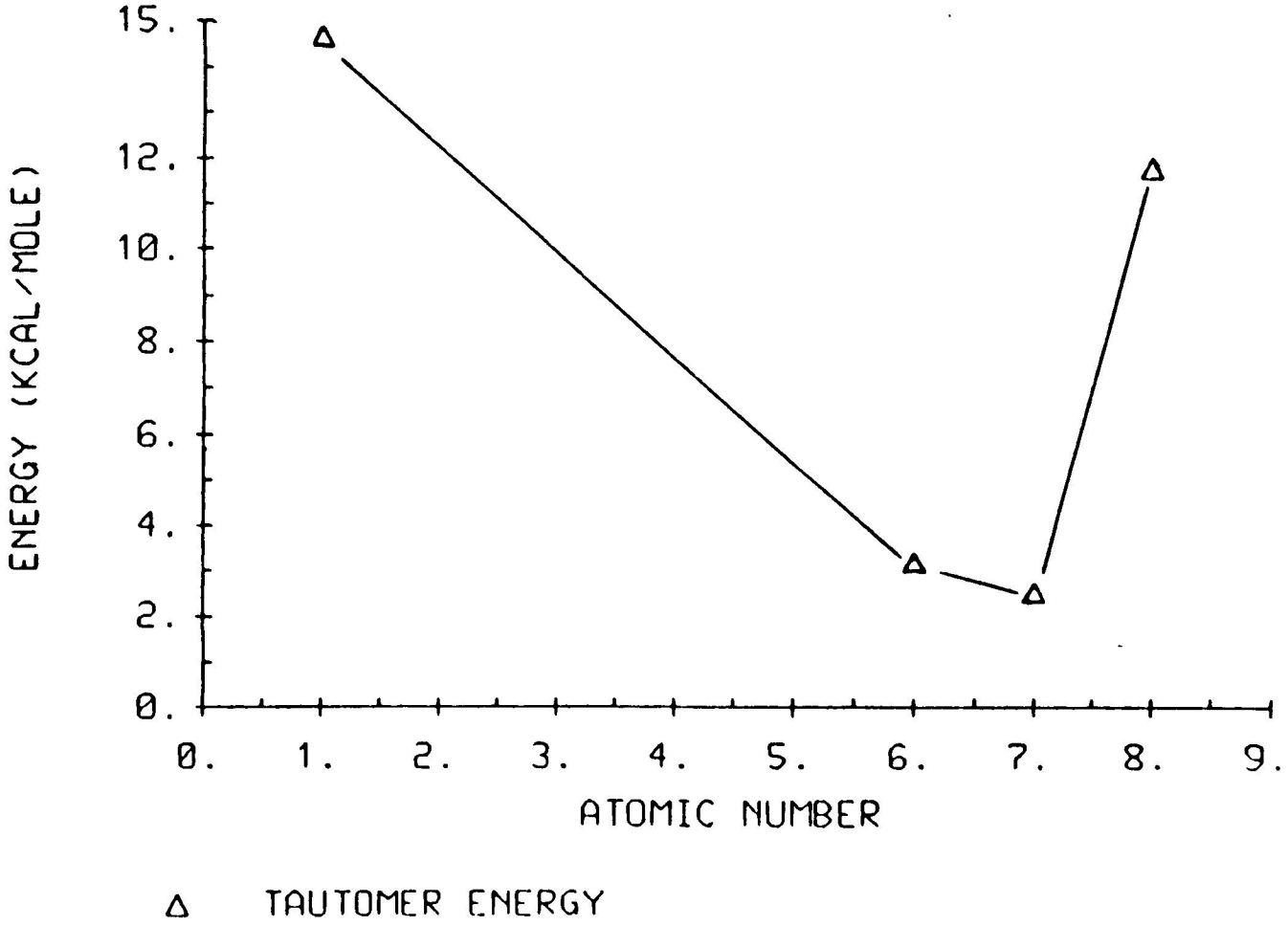



Figure 2-4. Relative tautomer stability vs. atomic number of the central atom of the substituent at $N_4$.

TABLE 2-3. Hydrogen Bonding Energies[a]

| Base Pair | $\Delta E^b$(STO-3G) | $\Delta E^b/HB^c$ |
|---|---|---|
| G:$C^{NH_2}$ | -35.26 | -11.75 |
| A:$C^{OH}$ | -34.73 | -17.36 |
| G:C | -29.99 | -9.9 |
| G:$C^{OH}$ | -28.99 | -9.6 |
| A:C | -19.40 | -9.7 |
| A:T | -13.64 | -6.8 |
| A:U | -12.07 | -6.0 |
| A:$C^{NH_2}$ | -5.52 | -2.7 |

[a]Energies are in kcal/mole.
[b]$\Delta E = E_{base\ pair} - (E_X + E_X)$
X = appropriate nucleic acid base
[c]HB = Hydrogen bond
C = Cytosine
$C^{OH}$ = $N^4$-hydroxycytosine
$C^{NH_2}$ = $N^4$-aminocytosine
$C^{CH_3}$ = $N^4$-methylcytosine

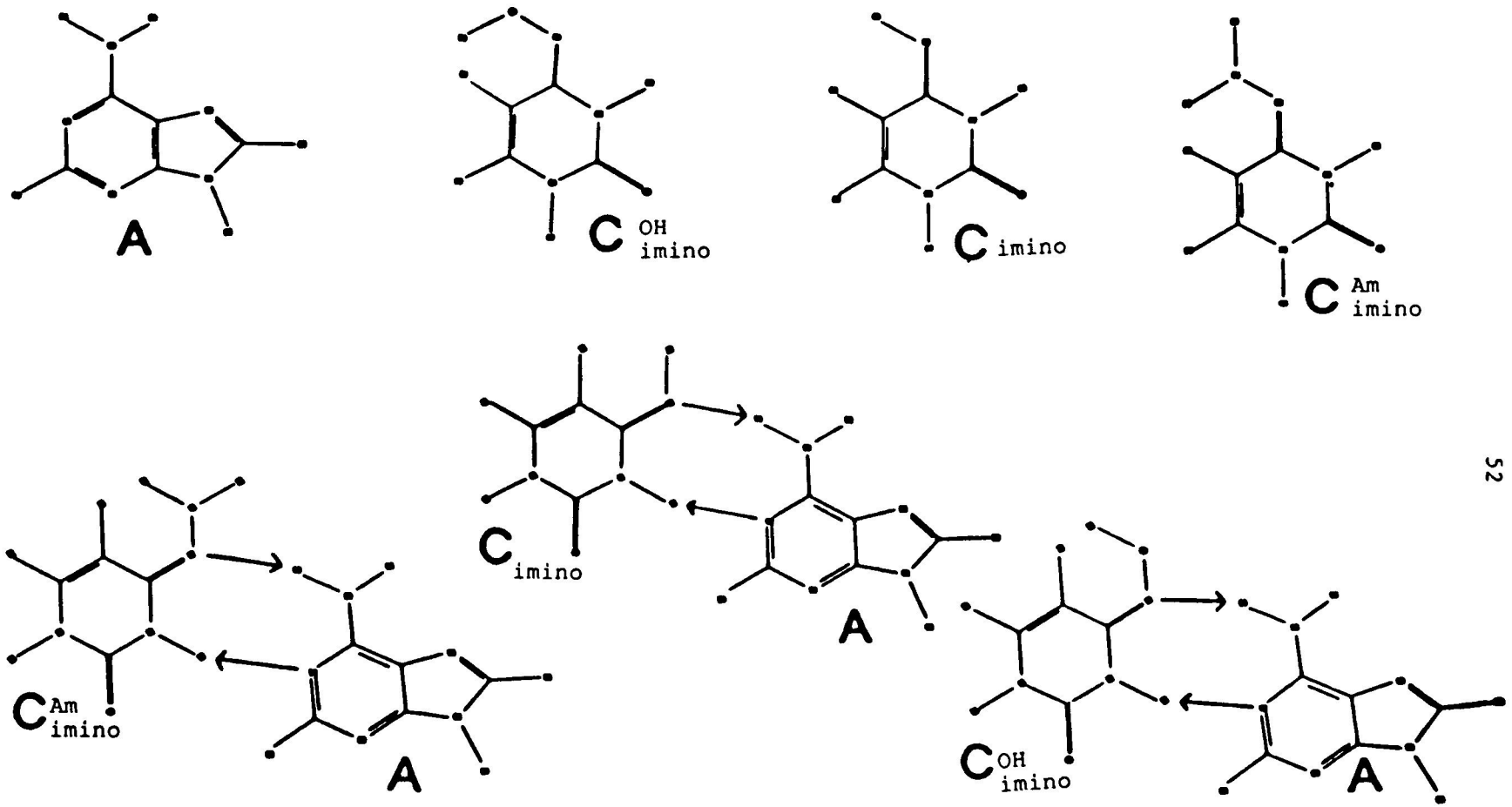





Figure 2-5. Localized molecular orbitals for A, $C^X$(imino) and A:$C^X$(imino) and their H-bonds.

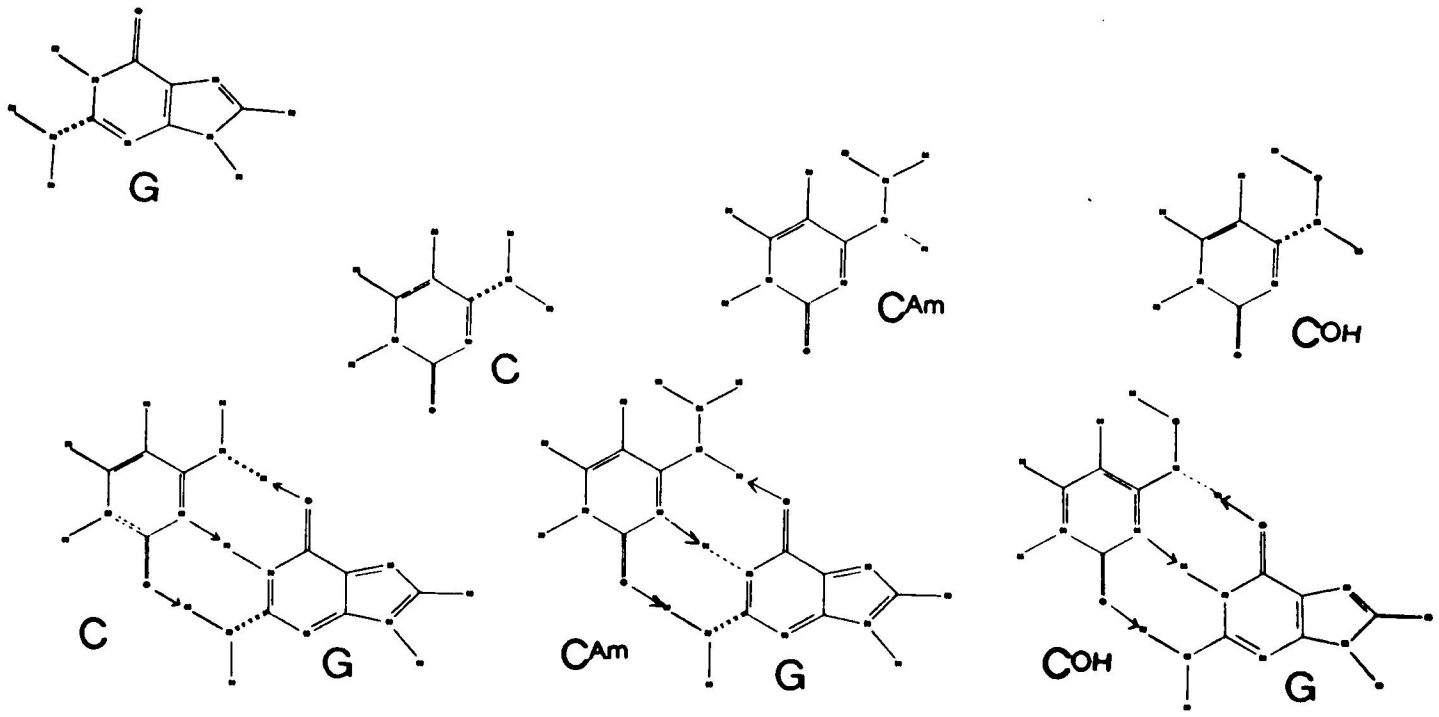



Figure 2-6. Localized molecular orbitals for G, $C^X$, and G:$C^X$, and their H-bonds.

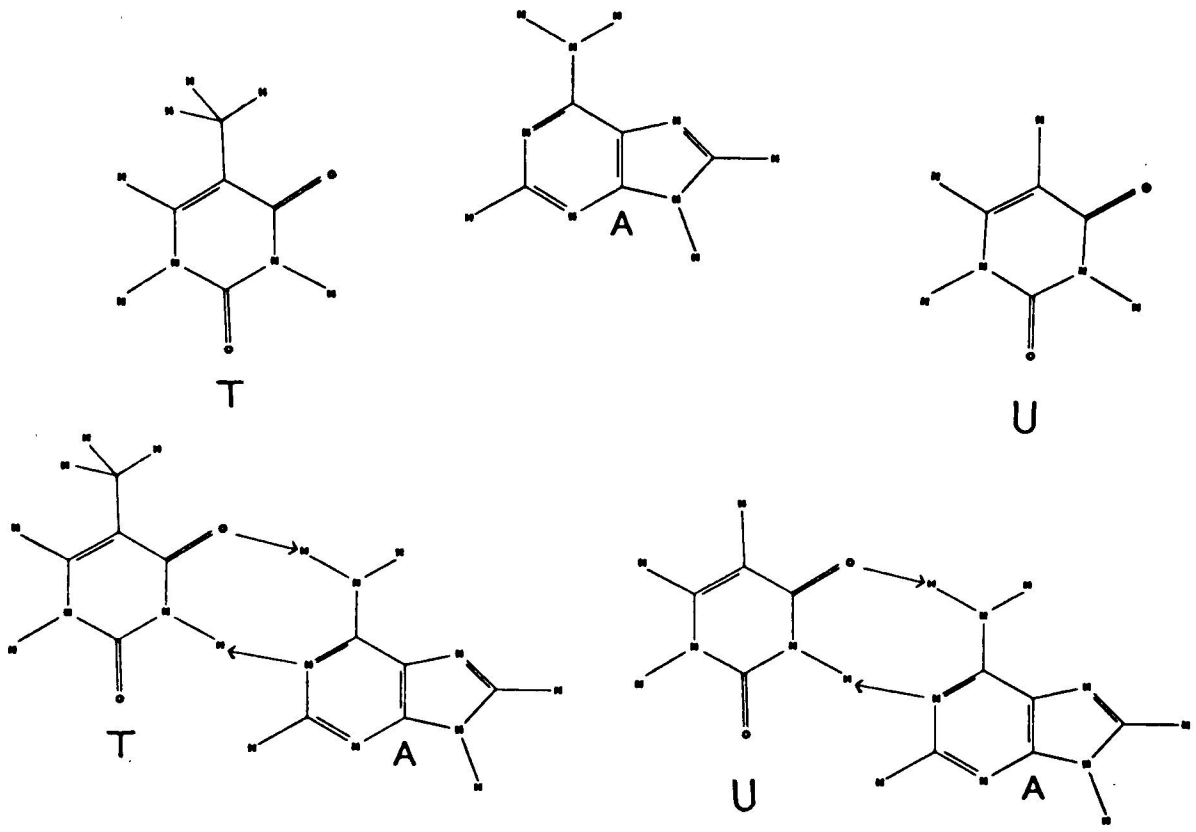



Figure 2-7. Localized molecular orbitals for A, T, U, A:T, and A:U, and their H-bonds.

relative stability of the A:$C^X$ pairs (Table 2-3, Figure 2-5) show that the order of stability is the reverse of the sequence found for the G:$C^X$ pairs (Table 2-3, Figure 2-6). The order of stability for the G:$C^X$ and A:$C^X$ pairs was calculated to be:

$$G:C^{Am} > G:C > G:C^{OH} \quad (1),$$

$$A:C^{OH} > A:C > A:C^{Am} \quad (2),$$

and the overall order of pair stability is:

$$G:C^{Am} > A:C^{OH} > G:C > G:C^{OH} > A:C > A:T > A:U > A:C^{Am} \quad (3).$$

The G:C pair is predicted to be more stable than the A:T pair (Table 2-3, Figure 2-7), in agreement with the *ab-initio* nonempirical calculations of Hozba and Sandorfy.[23] Experimental results show that the A:T regions of the helix are the first to unwind upon heating, and that the A:T content of the helix affects the melting temperature of DNA,[2] indicating that the A:T is not as strong as the G:C base pair.

The reaction schemes proposed by Negishi et.al.[6] for the A:T ---> G:C, and the G:C ---> A:T transitions caused by $C^{Am}$ and $C^{OH}$ are:

$$A:T \dashrightarrow A:C^{Am} \dashrightarrow G:C^{Am} \dashrightarrow G:C \quad (4)$$

and

$$G:C \dashrightarrow G:C^{OH} \dashrightarrow A:C^{OH} \dashrightarrow A:T \quad (5).$$

The broken arrows represent a series of biological reactions associated with the replication cycle. In the first sequence of reaction (4), the least stable pair (A:$C^{Am}$) is

formed. The imino tautomer of $C^X$ (II) is the one that pairs with adenine; however, in the G:$C^{Am}$ pair the amino tautomer of $C^X$ (I) pairs with guanine. Thus, the A:$C^{Am}$ pair may be formed in preference to the more stable G:$C^{Am}$ pair due to the low relative energy for tautomer II of $C^{Am}$ ($\Delta E_{I-II}$ = 2.5 kcal/mole). Once this step occurs, the reaction proceeds as written. The final step of reaction (4) occurs despite the relative stability of the G:$C^{Am}$ complexes because the rate of incorporation of $C^{Am}$ into DNA is not as fast as the rate of incorporation of C into DNA.[6] The analogy used to explain reaction scheme (4) may also be used to explain reaction scheme (5). The $\Delta E_{I-II}$ for $C^{OH}$ is higher than the $\Delta E_{I-II}$ for C, so $C^{OH}$ is not expected to tautomerize as readily as $C^{Am}$. The high $\Delta E_{I-II}$ for $C^{OH}$ allows tautomer I of $C^{OH}$ to pair with G as opposed to tautomer II of $C^{OH}$ pairing with A. Thus, the overall results indicate that both of the proposed reaction schemes (4 or 5) may occur.

Localized Molecular Orbitals for Base Pairs

Base Pairs with Adenine

The LMO's for A and A:$C^X$ are shown in Figure 2-5. The LMO centers having an electron population density (EPD) < 0.35 e are considered as nonbonded LMO's. The LMO centers having 0.35 to 0.60 e are drawn as dotted lines and those having EPD > 0.60 e are drawn as solid lines.[24] The LMO's for the A:$C^X$ pairs show that the H atoms which are involved in H-bonding undergo a decrease in electron density (an

average of 0.10 e) upon pairing. A comparison of the A:$C^X$ (Figure 2-5), A:T and A:U pairs (Figure 2-7) does not reveal any significant differences between the O---H-N, and N---H-N H-bonding interactions. The double bonds in the LMO scheme for A in the A:$C^X$ pairs are in different positions than the double bonds in the LMO scheme for A alone (Figure 2-5). In contrast, the LMO's of A do not show any shifting of $\pi$ electron density when paired with uracil or thymine (Figure 2-7). Thus, the change in $\pi$ bond arrangements occurring in A upon pairing with C may be an indication that A stabilizes tautomer II of C. Additional support for these findings is derived from the NMR results of Stolarski et.al.[25a] which show an increase in the amino/imino tautomeric exchange rate with an increase in concentration of the base which binds to the imino tautomer. Our results therefore suggest a possible explanation for the observed order of stability for the H-bonded complexes of adenine with the $N^4$-substituted cytosine series. This explanation introduces a concept which has been defined as the intrinsic stability.[20] The more intrinsically stable a structure, the less interaction energy it will give upon pairing (solvation, etc.). In this study, the $\Delta E_{I-II}$'s are used as a measure of the intrinsic stability of these compounds. Since $\Delta E_{I-II}$ for $C^{OH}$ is less than $\Delta E_{I-II}$ for C, the A:$C_{(imino)}$ pair should be more stable than the A:$C^{OH}_{(imino)}$ pair due to the higher $\Delta E_{I-II}$ for C; however, this result is not predicted by our calculations.

Therefore, some additional factors may be involved in determining the relative stability of these H-bonded systems. A hydroxyl group can increase the proton affinity of its nearest neighbor.[26] Mulliken population analysis shows a decrease in electron density (0.46 e) in the $p_z$ orbital of the imino nitrogen of C upon pairing with adenine, while the $p_z$ orbital of the imino nitrogen of $C^{OH}$ shows a larger decrease in electron density (0.48 e) upon pairing with adenine. Thus, the interaction induced by the presence of the OH group on the imino nitrogen of $C^{OH}_{(imino)}$ may cause the $A:C^{OH}_{(imino)}$ pair to be more stable than the $A:C_{(imino)}$ pair.

The $\Delta E_{I-II}$ for $C^{Am}$ is very low. Therefore, $C^{Am}_{(imino)}$ should not have a large stabilization effect upon H-bonding with adenine. This is consistent with our results which show that the $A:C^{Am}_{(imino)}$ forms the least stable base pair in the adenine series.

Base Pairs with Guanine

The LMO's for G, $C^X$, G:C, $G:C^{Am}$, $G:C^{OH}$ are shown in Figure 2-6. The LMO structure of C (Figure 2-6) shows two partial double bonds between $C_4$ and the external nitrogen, but when C is paired with G, two partial bonds are located between the $C_4$ nitrogen and the H-bonding hydrogen and between $N_1$ and $C_2$ in C. The LMO structure of G in the G:C base pair shows a partial double bond between $C_2$ and the external nitrogen. The $C_4C_5$, $C_2N_3$, and $N_7C_8$ double bonds of

G shift to the $C_5N_7$, $N_3C_4$ and $C_8N_9$ positions when paired with C.

The LMO structure of $C^{Am}$ (Figure 2-6) shows a partial and a full bond between $C_4$ and the external nitrogen. In the G:$C^{AM}$ pair, there is only one bond between $C_4$ and the $NH_2$ group of $C^{Am}$. The hydrogen atom at the $N_1$ position of G in the G:$C^{Am}$ pair has only a partial bond. This suggests that tautomerization of $C^{Am}$ may cause deprotonation of guanine in the double helix. Other studies have also indicated that tautomerization can occur due to an exchange of protons in H-bonding interactions of substituted nucleic acids.[25]

The LMO structure of $C^{OH}$ (Figure 2-6) shows two partial bonds between $C_4$ and the external nitrogen. When $C^{OH}$ is paired with G, only one full bond is present between $C_4$ and the external nitrogen, and a partial bond is located between the external nitrogen and the hydrogen that participates in hydrogen bonding. A comparison of the LMO results for $C^{OH}$, $C^{Am}$, and the G:$C^{Am}$ and G:$C^{OH}$ base pairs indicates that the presence of the hydroxyl group increases the acidity of the amino hydrogen as indicated by the numbers of partial and full bonds present between the six member ring and the external nitrogen (Figure 2-6). The LMO's for paired $C^{OH}$ show a three center bond between $N_5$, $C_4$ and $N_4$. The amount of electron density between $C_4$ and $N_4$ (0.26 e) is below the amount required to be considered bonded (0.35 e), so the LMO

structure for G:$C^{OH}$ (according to the criterion prescribed by Kleier and Lipscomb[24]) is drawn with five bonds at $C_4$. An increase in bonding electron density of 0.55 e at $C_4$ occurs upon formation of the G:$C^{OH}$ pair.

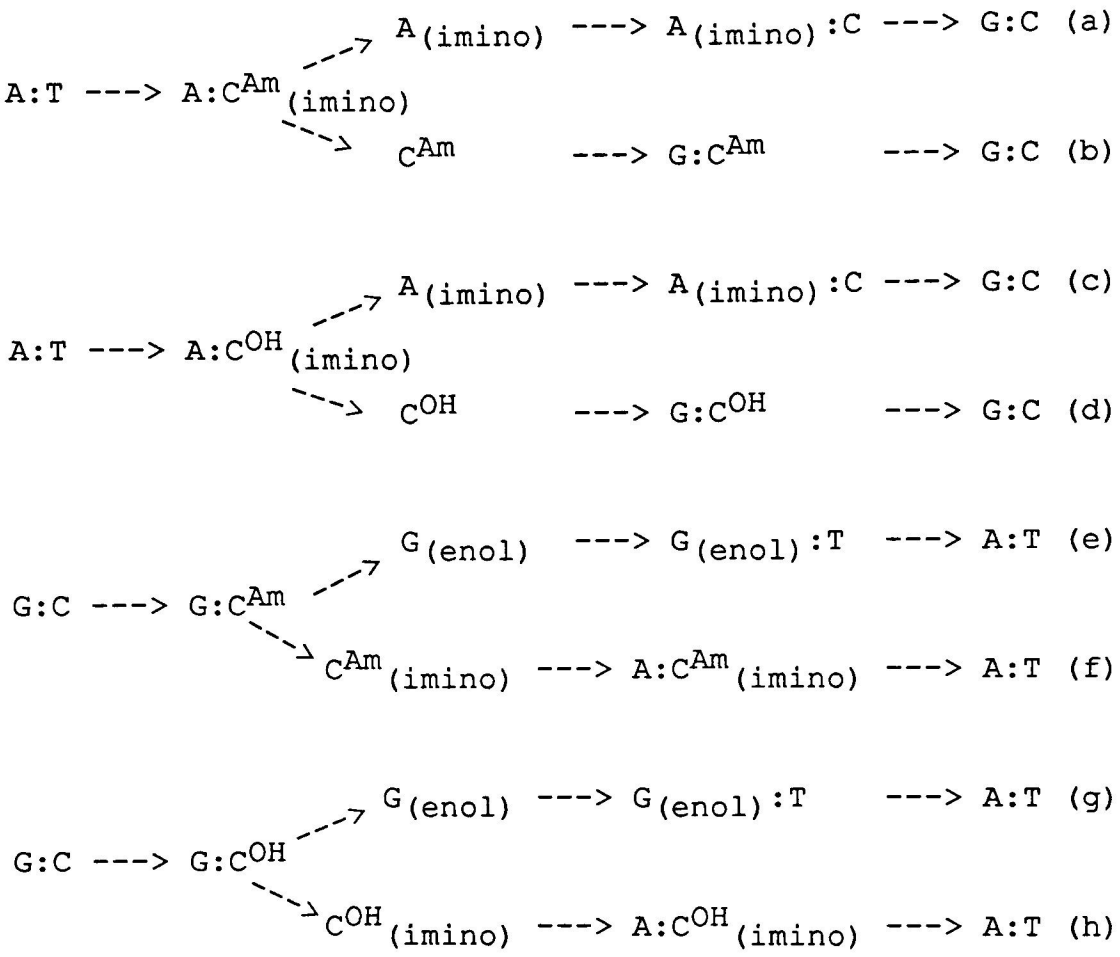
The LMO results for the G:$C^{X}$ series offer a possible explanation for the calculated order of stability of the G:$C^{X}$ pairs. In the G:$C^{X}$ pairing series (scheme 5), G pairs with tautomer I of C. The $\Delta E_{I-II}$ for C is large (14.65 kcal/mole), so according to the intrinsic stability concept, there should be a minimum of interaction energy between G and C because both of these structures are in their lowest energy state. The G:$C^{OH}$ pair is the least stable pair (Table 2-3). The inductive effect of the hydroxyl group in $C^{OH}$ causes the imino nitrogen to hold its electrons more tightly. Since these electrons are held tightly, there is an increase in the acidity of the hydrogen participating in H-bonding with G due to a weakening of the N-H bond. The weakened N-H bond in $C^{OH}$ can destabilize the pair, which may cause the G:$C^{OH}$ base pair to be the least stable in the G:$C^{X}$ series. Therefore, these results suggest that the effect of the OH group on imino nitrogens is different than its effect on amino nitrogens, and that when this pair separates, deprotonation of $C^{OH}$ may occur. The differences in the effect of the OH group on the amino and imino nitrogens can be due to the different charge on the amino (+) and imino (-) nitrogens.[27]

$N_4$-aminocytosine has a very low $\Delta E_{I-II}$, which is attributed to electron donation into the six member ring of cytosine. Thus, the amino substituent at $N_4$ seems to cause an increase in the amount of electron density donated to the ring. The low $\Delta E_{I-II}$ for $C^{Am}$ (2.5 kcal/mole) indicates a decrease in the potential barrier for tautomerization. The tendency for $C^{Am}$ to tautomerize is so great that it partially removes the H atom from the $N_1$ position of G in the G:$C^{Am}$ pair. The strong attraction for the $N_1$ hydrogen of G causes the G:$C^{Am}$ pair to be the most stable. This type of H-bonding interaction has been recently described by Huyskens, et al.[28b] Therefore, the order of stability of the G:$C^X$ pairs seems to be dependent on the tautomeric stability of $N_4$-substituted cytosine and the nearest neighbor effect of the substituent.[26]

Reaction Mechanism

Our results suggest that the substituents may affect the protonation/deprotonation properties in the H-bonding interaction of substituted nucleic acids via changes in tautomeric stabilities. Stolarski et al.[25a] also attribute the proton transfer between H-bonded nucleic acids to changes in tautomeric stabilities. A tautomerization occurring in either of the G:$C^X$ pairs would cause G to convert to the enol form while $C^X$ is converted to the imino form (similarly to the mechanism that occurs for the A:$C^X$ pairs). These results suggest that additional steps may be

occurring in the reaction mechanisms of these mutagenic compounds. The possible pathways are as follow:




Pathways (a) through (d) represent the A:T ---> G:C transition with $C^{Am}$ and $C^{OH}$, whereas pathways (e) through (h) represent the G:C ---> A:T transitions with $C^{Am}$ and $C^{OH}$. The initiation step is believed to be dependent upon the tautomerization of $C^{X}$. Since $\Delta E_{I-II}$ for $C^{Am}$ is lower than $\Delta E_{I-II}$ for $C^{OH}$ ($\Delta E_{I-II}$ = 11.7 and 2.5 kcal/mole for $C^{OH}$ and $C^{Am}$, respectively), the A:T ---> G:C mechanism should occur via pathways (a) and (b) instead of via pathways (c) and (d). The experimental studies with $C^{Am}$ were not designed to

detect the G:C ---> A:T transition,[6-8] so more experimental data is needed to show whether $C^{Am}$ is capable of inducing the G:C ---> A:T transition. Pathways (e) and (f) describe the mechanism that may be operative when $C^{Am}$ causes the G:C ---> A:T transition. Support for step 2 in reactions (e) and (f) comes from our LMO calculations which indicate that the $N_1$ position of G is partially deprotonated when paired with $C^{Am}$ (Figure 2-6). We interpret this to be an indication that upon separation of the pair, guanine may be deprotonated.

The pathways for the G:$C^X$ pairs are analogous to the pathways for A:$C^X$ pairs, so the $\Delta E_{I-II}$'s for the $C^X$'s are believed to play an important role in determining the initiation step of the mutagenic pathway for the guanine series. The G:C ---> A:T transition for $C^{OH}$ is described by pathways (g) and (h). The $\Delta E_{I-II}$'s for $C^{Am}$ and $C^{OH}$ suggest that the (g) and (h) sequence should occur in preference to the (e) and (f) sequence. Pathways (c) and (d) describe the A:T ---> G:C transition for $C^{OH}$.[9] Our calculations support experimental results which show that in a specific cell type, $C^{OH}$ and $C^{Am}$ should cause opposing transitions.[6-8] If tautomerization is the key to understanding these transitions, then opposing transitions for $C^{Am}$ and $C^{OH}$ should occur no matter which biological species or experimental parameters are used.

Biological Implications

It has been shown that specific conformational and electronic changes may be induced in DNA by various agents (ie. protonation or deprotonation,[29] alkylation,[30] humidity,[31c,32] metal ions,[31a,b,33] etc.). Specific biological responses are associated with the generation of proton gradients, opening of ion channels, hormone production, enzyme activation (and synthesis, etc.). The peptide sequences, proton gradients and ions bind to enzymes and DNA to induce specific conformational changes. These conformational changes are responsible for the biological activity of these macromolecular polymers. By inducing specific electronic and conformational changes in the nucleic acid helices through attachment of a specific group at strategic locations, it should be possible to initiate cellular reactions (eg., a cell differentiation cycle).

This should be possible by manipulating the mRNA's or the DNA region responsible for their synthesis. The modification of individual nucleic acid bases by an invading organism may label an essential binding site where one may not otherwise exist by inducing site specific transitions or by binding to a region with substituted nucleic acids. (This result has not been detected experimentally probably due to the complexity of the mechanism involved, so this suggestion is speculative.) Once the viral genetic material binds to the host DNA (or RNA), the transformation process of the

cell (to a cancerous one) can be completed. By considering some of the information available on mutations and cancer inducing mechanisms, it seems that the best method for attacking a cancer causing cell should be similar to the method used by the cancerous cell to attack the host. (This suggestion is based on the assumption that the method of cellular invasion is known). An alternative approach or a method which would be useful in controlling metastasis is by designing a pseudo binding site for the invading cell which would take it into a degradative cycle.

# CONCLUSIONS

*Ab-initio* SCF calculations on cytosine, adenine, guanine, their tautomers using (PRDDO, STO-3G, and 4-31G basis sets) and H-bonded base pairs (using STO-3G basis set) indicate that the substituted group can alter the relative tautomer stability by increasing electron donation into the ring, or by changing the acidity of a tautomeric hydrogen. When this tautomeric hydrogen participates in H-bonding, it may affect the stability of the pair.

Our results also suggest that tautomerization of the bases may occur during the replication cycle, and that the presence of tautomers can lead to mutations.

Tautomerization may also affect the stacking interactions between the individual bases in the helix due to the changes in the $\pi$ electron system. Therefore, strategic manipulation of H-bonding and $\pi$-$\pi$ interactions may allow one to induce a desired cellular response by using the proper combination of point mutations, site specific intercalators (or binders), and biological communication signals.[9,40-43] These signals induce biologically active conformations in DNA and RNA and their fragments. Proper development of these procedures should provide additional methods for treatment of individuals with cancer, and may also be a starting point for the development of new methods for combating viral and bacterial infections.

REFERENCES


1. Freese, E.; Proc. Nat. Acad. Sci. USA **1959**, 45, 622.

2. Saenger, W. "Principles of Nucleic Acid Structure", Springer-Verlag:New York, N. Y., 1984; Chapter 6.

3. (a) Negishi, K.; Harada, C.; Ohara, Y.; Oohara, K.; Nitta, N.; Hayatsu, H.; Nuc. Acids Res. **1983**, 11, 5223.

   (b) Popowska, E.; Janion, C.; Nuc. Acids Res. **1975**, 2, 1143.

   (c) Pietrzykowska, I.; Mut. Res. **1973**, 19, 1.

   (d) Ref. 2, Chapter 7.

4. (a) Hillen, W.; Egert, E.; Lindner, H.; Gassen, H.; Vorbruggen, H.; Mullerstr, S.; J. Carb. Nucleos. Nucleot. **1978**, 5, 23.

   (b) Ref. 2, p. 114.

5. (a) Janion, C. Mutation Research **1978**, 56, 225.

   (b) Chu, B.; Brown, D.; Burdon, M. Mutation Research **1974**, 23, 267.

6. (a) Negishi, K.; Takahashi, M.; Yamashita, Y.; Nishizawa, M.; Hayatsu, H.; Biochem. **1985**, 24, 7273.

   (b) Ref. 4.

   (c) Budowsky, E.; Sverdlov, E.; Spasokukotskaya, T.; Biochim. Biophys. Acta **1972**, 287, 195.

7. (a) Popowska, E.; Janion, C.; Biochem. Biophys. Res. Comm. **1974**, 56(2), 459.

   (b) Tessman, I.; Poddar, R.; Kumar, S.; J. Mol. Biol. **1964**, 9, 352.

8. Flavell, R.; Sabo, D.; Bandle, E.; Weissmann, C. J. Mol. Biol. **1974**, 89, 255.

9. Muller, W.; Weber, H.; Meyer, F.; Weissmann, C.; J. Mol. Biol. **1978**, 124, 343.

10. (a) Walker, G. A.; Bhatia, S. C.; Hall, Jr., J. H. J. Am. Chem. Soc. **1987**, 109, 7634.

(b) Walker, G. A.; Bhatia, S. C.; Hall, Jr., J. H. J. Am. Chem. Soc. **1987**, 109, 7629.

(c) Jones, P. Y.; Bhatia, S. C.; Hall, Jr., J. H., to be submitted to J. Am. Chem. Soc.

(d) Pambi, D. M.S. Thesis, Atlanta University 1988, Atlanta, Ga.

11. Roothaan, C.; Rev. Mod. Phys. **1951**, 23, 69.

12. Del Bene, J.; J. Phys. Chem. **1983**, 87, 367.

13. Ohta, Y.; Tanaka, H.; Baba, Y.; Kagemoto, A.; J. Phys. Chem. **1986**, 90, 4438.

14. Gauss80, Quantum Chemistry Program Exchange #421.

15. Hehre, W.; Stewart, R.; Pople, J.; J. Chem. Phys. **1969**, 51, 2657.

16. Ditchfield, R.; Hehre, W.; Pople, J.; J. Chem. Phys. **1971**, 54, 724.

17. (a) Boys, S.; Rev. Mod. Phys. **1960**, 32, 306.

(b) Foster, J.; Boys, S.; Rev. Mod. Phys. **1960**, 32, 300.

(c) Boys, S. "Quantum Theory of Atoms, Molecules, and the Solid State"; Per-Olov Lowdin, Ed.; Academic Press: New York, N. Y., 1966; p. 253.

18. Halgren, T.; Lipscomb, W.; J. Chem. Phys. **1973**, 58, 1569.

19. Schlegel, H.; Gund, P.; Fluder, E. J. Am. Chem. Soc. **1982**, 104, 5347.

20. Bernasconi, C; Acc. Chem. Res. **1987**, 20, 301.

21. (a) Marcus, R. Annu. Rev. Phys. Chem. **1964**, 15, 155.

(b) Marcus, R. J. Phys. Chem. **1968**, 72, 891.

(c) Cohen, A.; Marcus, R. J. Phys. Chem. **1968**, 72, 4249.

22. Cieplak, P.; Bash, P.; Singh, C.; Kollman, P. J. Am. Chem. Soc. **1987**, 109, 6283.

23. Hozba, P.; Sandorfy, C.; J. Am. Chem. Soc. **1987**, 109, 1302.

24. Kleier, D.; Lipscomb, W.; International Journal of Quantum Chemistry **1977**, S4, 73.

25. (a) Stolarski, R.; Kierdaszuk, B.; Hagaberg, C.; Shugar, D.; Biochem. **1987**, 26, 4332.

(b) Iwahashi, H.; Kyogoku, Y.; J. Am. Chem. Soc. **1980**, 102, 2913.

26. Scheiner, S.; Acc. Chem. Res. **1985**, 18, 174.

27. Scanlan, M.; Hillier, I. J. Am. Chem. Soc. **1984**, 106, 3737.

28. (a) Wilson, E.; Smith, Z.; Acc. Chem. Res. **1987**, 20, 257.

(b) Huyskens, P.; Haulait-Prison, Cl.; Collaer, H.; Prison, D.; J. Phys. Chem. **1987**, 91, 4860.

(c) Legon, A.; Millen, D.; Acc. Chem. Res. **1987**, 20, 39.

29. Sowers, L.; Fazakerley, V.; Kim, H.; Dalton, L.; Goodman, M. Biochem. **1986**, 25, 3983.

30. (a) Rhoads, R.; Hellmann, G.; Remy, P.; Ebel, J.-P. Biochem. **1983**, 22, 6084.

(b) Pramanik, P.; Kan, L. Biochem. **1987**, 26, 3807.

(c) Darzynkiewicz, E.; Ekiel, I.; Tahara, S.; Seliger, L.; Shatkin, A. Biochem. **1985**, 24, 1701.

31. (a) Laundon, C.; Griffith, J.; Biochemistry **1987**, 26, 3759.

(b) Koo, H.; Crothers, D.; Biochemistry **1987**, 26, 3745.

(c) Taillandier, E.; Ridoux, J.-P.; Liquier, J.; Leupin, W.; Denny, W.; Wang, Y.; Thomas, G.; Peticolas, W. Biochem. **1987**, 26, 3361.

32. (a) Saenger, W.; Hunter, W.; Kennard, O. Nature (London) **1986**, 324, 385.

(b) Kopa, M.; Fratini, A.; Drew, H.; Dickerson, R. J. Mol. Biol. **1983**, 163, 129.

33. Barr, R.; Pinnavaia, T. J. Phys. Chem. **1986**, 90, 328.

34. Ref. 2, Chapters 11 and 12.

35. Kuwajima, S.; Soos, Z. J. Am. Chem. Soc. **1987**, 109, 107.

36. Ishida, T.; Doi, M.; Ueda, H.; Inoue, M.; Scheldrick, G. J. Am. Chem. Soc. **1988**, 110, 2286.

37. Harvey, R.; Geacintov, N. Acc. Chem. Res. **1988**, 21, 66.

38. C. Pullman and B. Pullman Comptes rendus **1965**, 261, 2521.

39. Ref. 2, p. 315.

40. Wieringa, B.; Meyer, F.; Reiser, J.; Weissmann, C.; Nature **1983**, 301, 38.

41. (a) Quigley, G.;Ughetto, G.; Marel, G.; Boom, J.; Wang, A.; Rich, A.; Science **1986**, 232, 1255.

    (b) Wilson, W.; Wang, Y.; Chandrasekaran, K.; Yang, N.; Boykin, D. J. Am. Chem. Soc. **1985**, 107, 4989.

42. (a) Jordan, F.; Sostman, H.; J. Am. Chem. Soc. **1973**, 95, 6544.

    (b) Ref. 2, p.

43. Moser, H.; Dervan, P. Science **1987**, 238, 645.

# Chapter 3

## A Self-Consistent-Field-Study of Base Pairing for the Tautomeric Forms of Guanine and Cytosine

## Introduction

It is widely accepted that the DNA molecule exists in a double helix where the two polynucleotide chains are connected by complementary base pairs via hydrogen bonding. Although Watson and Crick postulated this theory, they also recognized that if base pairing adhered to such a rigid model, life on this planet would not display the vast variety of species.[1] The stability of base pairing is essential for the fidelity of genetic information and is delicately controlled by the protein-DNA interaction during transcription and replication of DNA.[2]

In DNA replication, two types of spontaneous substitution mutations can occur: transitions and transversions. When a transition occurs, a purine is replaced by another purine, or a pyrimidine is replaced by another pyrimidine; while in transversions, a purine is exchanged for a pyrimidine or vice versa.[3] This change is also generally accompanied by an anti to syn base rotation and a C-3'-endo to C-2'-endo change in sugar conformation.[4] These types of conformational changes in nucleotides are believed to be responsible for macromolecular changes (e.g., B --> Z DNA conformational change).[4] The formation of non-Watson-Crick base pairs in biomolecules can result in mutations. This also brings up the question of whether these mutations would eventually turn out to be a binding site for an invading cell. Watson and Crick[5] also suggested

that spontaneous mutations might be due to the occurrence of the tautomeric form of DNA bases which give rise to non-Watson-Crick base pairing. Drake et al.[6] observed the spontaneous mutation rate in bacteriophage T4 to be $10^{-8}$ to $10^{-10}$ mutations per newly synthesized DNA. The geometric structures of the non-Watson-Crick base pair and the Watson-Crick base pair are very close. The closeness of structures of different base pairs is such that they are not detected as misplaced and are passed by the enzymatic proofreading step during DNA replication.[7] This sequence of mutated DNA may be involved in a major biological reaction, and may be severe enough to cause loss of cell function or it may delete an essential binding site. Therefore, the protein that is synthesized by the mutated sequence may be adversely affected and will affect the resulting gene expression.[8]

Sarai et al.[2] investigated the effect of external hydrogen bonding interactions on the stability of Watson-Crick and non-Watson-Crick base pairing scheme for guanine-cytosine base pairs. In chapter 2, we focused on the effect of the subtituents on hydrogen bonding interactions in Watson-Crick and non-Watson-Crick base pairs. Tautomerization of nucleic acid bases within the helix may cause mispairing by reversing the directions of base pair hydrogen bonds.[2] These changes in base-base interactions along a helix may be necessary for DNA and RNA to adopt

biologically active conformations.

In chapter 3 we focus specifically on the hydrogen bonding interaction between the tautomers of guanine and cytosine to assist in understanding the role of the non-Watson-Crick base pairs in the formation of various forms (A, B, and Z) of DNA.

*Ab-initio* SCF calculations were performed to determine the relative stability of cytosine and its tautomers, as well as the tautomeric base pairs of guanine and cytosine. The biological implications of the SCF calculations are discussed.

## METHODS

*Ab-initio* SCF calculations were performed on the tautomeric forms of cytosine (3H, 1H, enol and imino), guanine (7H1H, 3H9H, 3H7H, imino and enol) and on the hydrogen bonding interaction for the tautomeric forms of guanine and cytosine. Optimized geometries for the DNA bases were taken from Del Bene's study of bases protonation.[9] The GAUSS86 program (QCPE No. 446)[10] with STO-3G basis set was employed for these calculations. The hydrogen bond distances were optimized with the STO-3G basis set (Table 3-1). The O6···H hydrogen bond was optimized while the N4--H bond distance was kept fixed for each tautomeric base pair investigated.

The numbering system for guanine (purine) and cytosine (pyrimidine) is shown in Figure 3-1.

TABLE 3-1. Optimized Hydrogen Bonding Geometries

| Tautomeric Base Pairs | Hydrogen Bond | R(X···Y)[a] | R(X···Y)[a,b] |
|---|---|---|---|
| C1H···G9H1H | N4-H···O6 | 2.54 | 2.66 |
| | N3···H-N1 | 2.72 | 2.71 |
| | O2···H-N2 | 2.93 | 2.77 |
| C1H···G7H1H | N4-H···O6 | 2.54 | |
| | N3···H-N1 | 2.63 | |
| | O2···H-N2 | 2.74 | |
| C1H···GIMINO | N4-H···O6 | 2.54 | |
| | N3···H-N1 | 2.63 | |
| | O2···H-N2 | 3.81 | |
| C3H···G3H7H | N4-H···O6 | 2.55 | |
| | N3···H-N1 | 2.79 | |
| | O2···H-N2 | 2.76 | |
| C3H···G3H9H | N4-H···O6 | 2.55 | |
| | N3···H-N1 | 2.79 | |
| | O2···H-N2 | 2.76 | |
| C3H···GENOL | N4-H···O6 | 2.55 | |
| | N3···H-N1 | 2.79 | |
| | O2···H-N2 | 2.76 | |
| CENOL···GIMINO | N4-H···O6 | 2.64 | |
| | N3···H-N1 | 2.63 | |
| | O2···H-N2 | 2.65 | |
| CENOL···G7H1H | N4-H···O6 | 2.64 | |

TABLE 3-1. (Continued)

| | | |
|---|---|---|
| | N3···H-N1 | 2.63 |
| | O2···H-N2 | 2.65 |
| CENOL···G9H1H | N4-H···O6 | 2.64 |
| | N3···H-N1 | 2.71 |
| | O2···H-N2 | 2.82 |
| C3···G9H1H | N4-H···O6 | 2.73 |
| | O2···H-N2 | 3.03 |
| C3···G7H1H | N4-H···O6 | 2.73 |
| | O2···H-N2 | 3.03 |
| C3···GIMINO | N4-H···O6 | 2.73 |
| | O2···H-N2 | 3.03 |
| C1H···GENOL | N4-H···O6 | 2.90 |
| | O2···H-N2 | 3.17 |
| C1H G3H9H | N4-H···O6 | 2.90 |
| | O2···H-N2 | 3.09 |
| C1H···G3H7H | N4-H···O6 | 2.90 |
| | O2···H-N2 | 3.09 |
| CENOL···GENOL | N4-H···O6 | 3.02 |
| | O2···H-N2 | 3.02 |
| CENOL···G3H9H | N4-H···O6 | 3.02 |
| | O2···H-N2 | 3.01 |
| CENOL···G3H7H | N4-H···O6 | 3.02 |
| | O2···H-N2 | 3.01 |

[a]Distance between heavy atoms in A
[b]Del Bene, J.E.; J. Mol. Struct. (Theochem), **1985**, 124, 201.

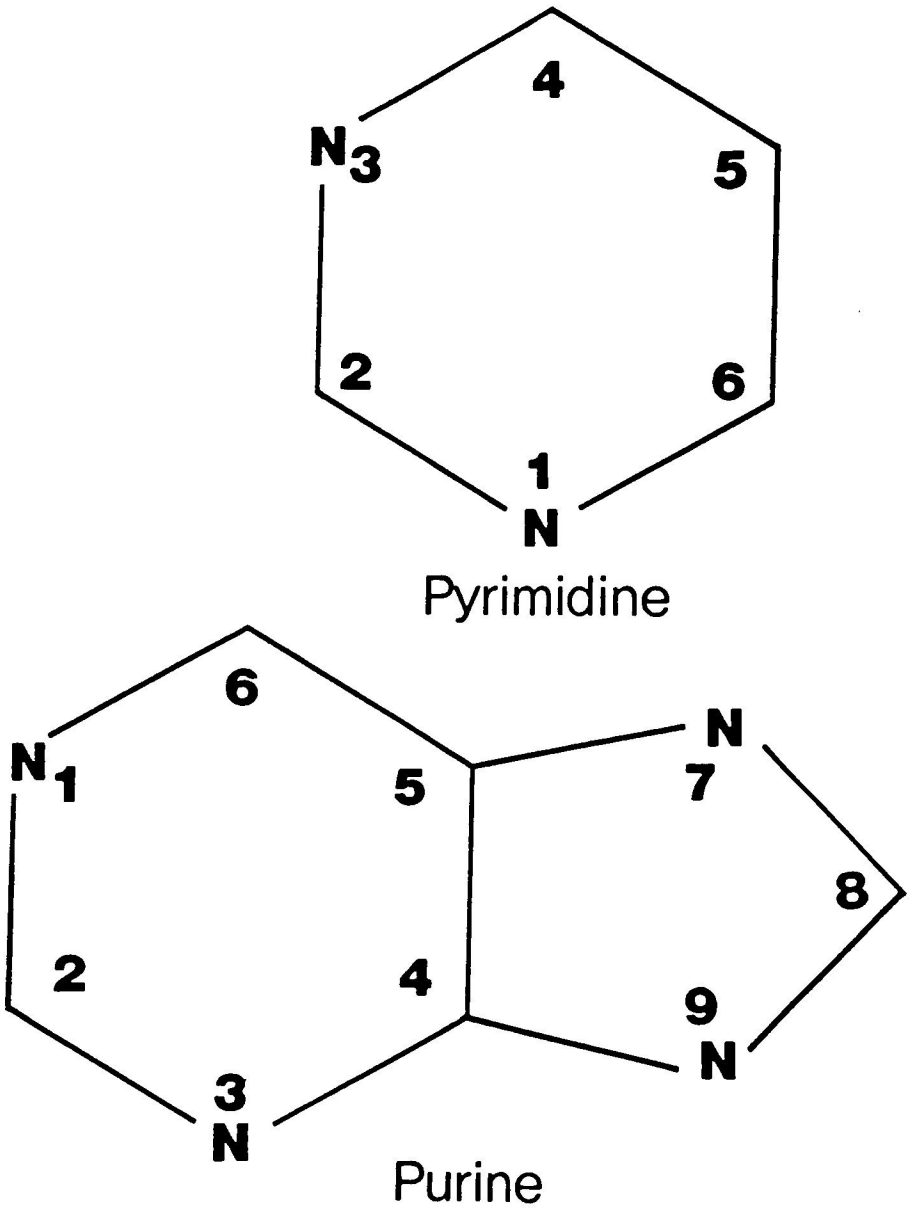
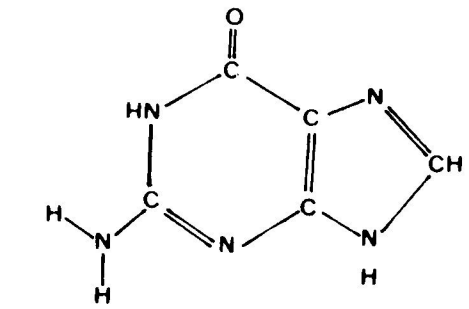



Figure 3-1. Numbering of atoms for purine and pyrimidine ring systems.

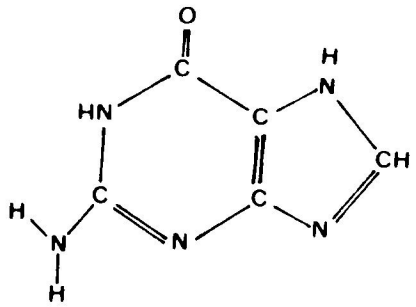
## Results and Discussion

The different tautomeric forms of guanine are shown in Figure 3-2, and the energy analysis is presented in Table 3-2. The relative energy ordering for the guanine tautomers is calculated to be 9H1H < 7H1H < imino < 3H7H < 3H9H < enol at STO-3G level.[11] Lee et al.[12] have experimentially observed significant formation of the enol guanine tautomers, but Miles et al.[13] and Wong et al.[14] concluded that the amount of enol tautomer present was negligible in their studies.

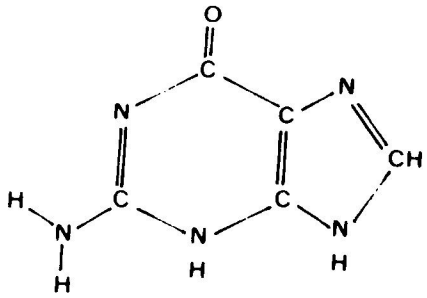
Lee et al[12] observed the tautomeric forms of cytosine to exist in a high percentage (15 =3%) at room temperture in neutral aqueous solution. The energy analysis for the tautomeric forms of cytosine is shown in Figure 3-3, and the energy analysis is presented in Table 3-3. The relative ordering at the STO-3G level calculations for the cytosine tautomers is predicted to be enol < C1H < imino(trans) < imino(cis) < C3H. The enol tautomer is more stable than the C1H tautomer by 5.39 kcal/mole. The relative ordering at the 3-21G level for the cytosine tautomers is calculated to be C1H < enol < imino (cis) < C3H < imino (trans). The C1H tautomer is more stable than the enol tautomer by 9.60 kcal/mole. Scanlan et al.[15] also have calculated the relative ordering at 3-21G level with STO-3G and 3-21G optimized geometries for the cytosine tautomer to be C1H < imino < enol < C3H. The C1H tautomer is predicted to be



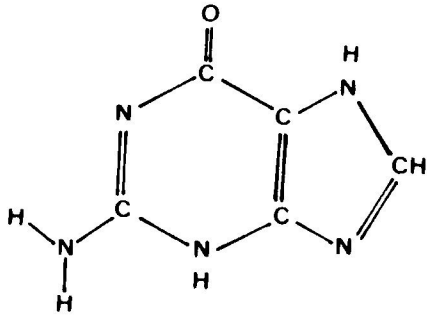
9-H, 1-H
I



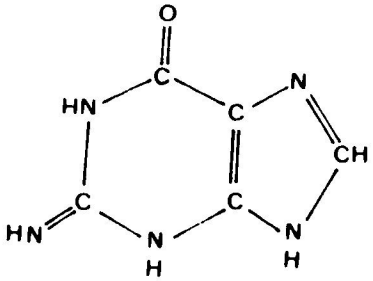
7-H, 1-H
II



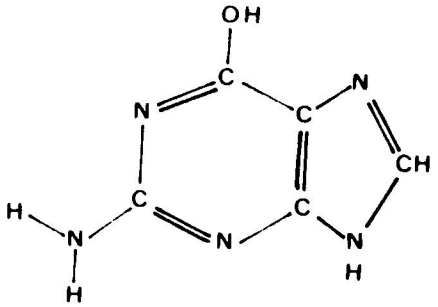
3-H, 9-H
III



3-H, 7-H
IV



IMIDO
V



ENOL
VI



Figure 3-2. Possible tautomers of guanine.

TABLE 3-2. Energy Analysis For The Tautomeric Forms of Guanine

| Tautomeric Forms | Basis Sets | | | | | |
|---|---|---|---|---|---|---|
| | $E^a$(PRDDO) | $\Delta E^b$ | $E^a$(STO-3G) | $\Delta E^b$ | $E^a$(4-31G) | $\Delta E^b$ |
| **9H,1H** | -537.9753 | 0 | -532.4613 | 0 | -538.5567 | 0 |
| **7H,1H** | -537.9529 | 14.49 | -532.4309 | 19.08 | -538.5392 | 10.99 |
| **3H,9H** | -537.9142 | 38.30 | -532.3941 | 42.17 | -538.5023 | 34.10 |
| **3H,7H** | -537.9084 | 41.97 | -532.3828 | 49.29 | -538.5067 | 31.36 |
| **IMINO** | -537.9120 | 39.72 | -532.3906 | 44.39 | -538.5112 | 28.53 |
| **ENOL** | -537.9300 | 28.42 | -532.4092 | 32.72 | -538.4969 | 37.53 |

[a]The total energies (E) in a.u.
[b]The relative energy ($\Delta E$) in kcal/mole

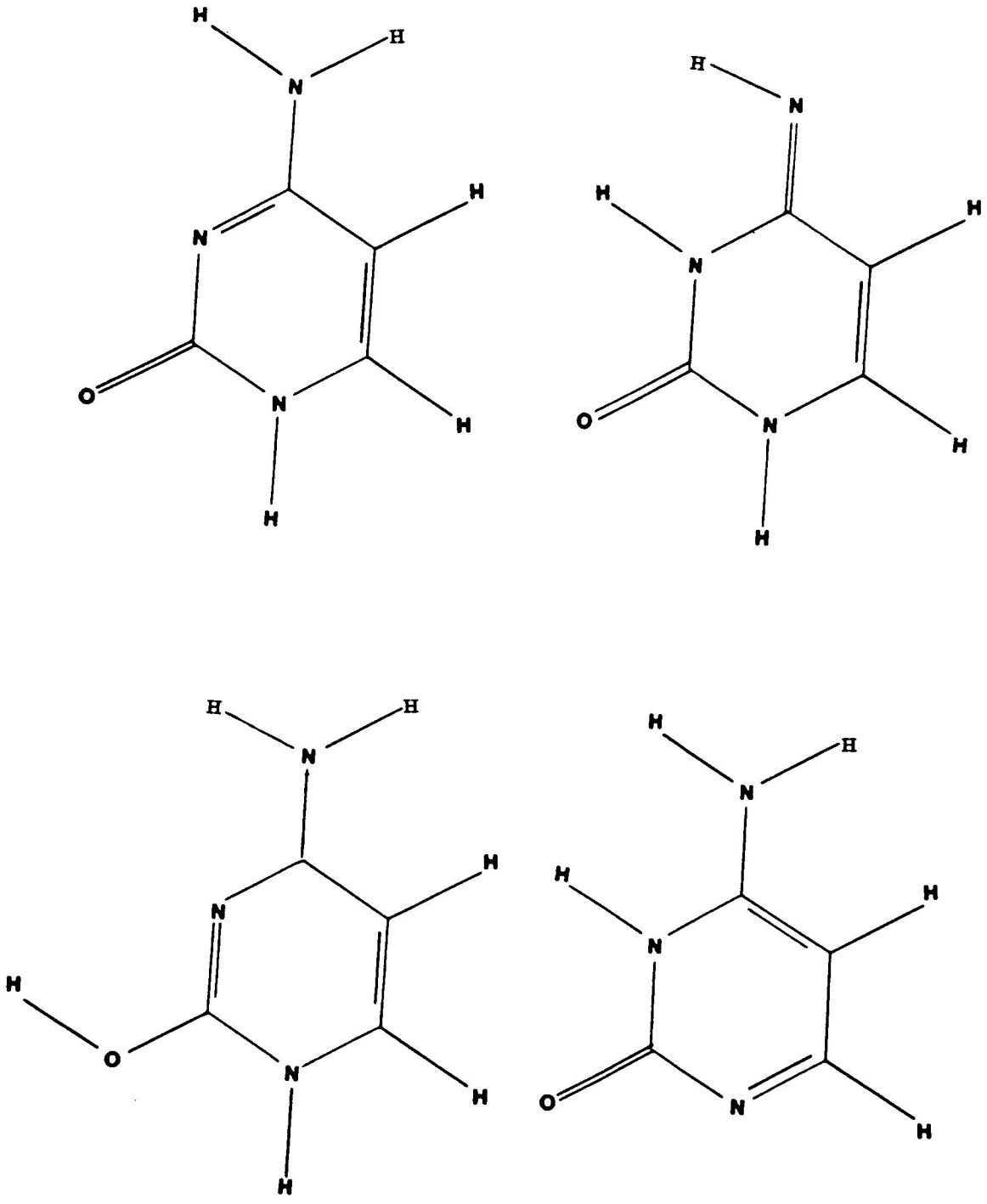


Figure 3-3. Possible tautomers of cytosine.

TABLE 3-3. Energy Analysis For The Tautomeric Forms of Cytosine

| Tautomeric Forms | Basis Sets | | | | | |
|---|---|---|---|---|---|---|
| | $E^a$ (STO-3G) | $\Delta E^b$ | $E^a$ (3-21G) | $\Delta E^b$ | $E^a$ (4-31G) | $\Delta E^b$ |
| **1H** | -387.5452 | 5.39 | -390.4058 | 0 | -392.0152 | 0 |
| **3H** | -387.5163 | 23.53 | -390.3835 | 13.99 | -391.9913 | 14.99 |
| **ENOL** | -387.5538 | 0 | -390.3905 | 9.60 | -391.9980 | 10.79 |
| **IMINO (cis)** | -387.5343 | 12.24 | -390.3889 | 10.60 | -391.9974 | 11.16 |
| **IMINO (Trans)** | -387.5395 | 8.97 | -390.3871 | 11.73 | -391.9941 | 13.24 |

[a]The total energies (E) in a.u.
[b]The relative energy ($\Delta E$) in kcal/mole

more stable than the imino tautomer by 7.5 kcal/mole. Calculations performed with an extended basis set (4-31G) using STO-3G optimized geometries are reported to give more reliable relative energies than those obtained at the minimal basis set level.[16] The relative ordering at 4-31G level for the tautomers is calculated to be C1H < enol < imino (cis) < imino (trans) < C3H. The imino (cis) tautomer was used in the hydrogen bonding study because it is predicted to be lower in energy than the imino (trans) tautomer by 2.12 kcal/mole. Moverover, the imino (trans) cannot participate in hydrogen bonding for the base pairs we have investigated. Nuclear magnetic resonance (NMR) study by Czerminski et al.[17] of cytosine in solution indicated that the C1H and C3H tautomers of cytosine were present, but the imino tautomers predominated in apolar solvents.[18] The NMR study[17] shows that the solvent effect plays a role in stablizing these tautomers.

Hydrogen bonds are mainly electrostatic in character, and they play a key role in stabilizating protein and nucleic acid secondary structure.[19] The eighteen possible base pairing schemes of guanine and cytosine are shown in Figure 3-4. The energy analysis of the types of hydrogen bonds for the tautomeric base pairs of guanine and cytosine is presented in Table 3-4. In our study, several of the tautomeric G-C base pairs have the usual three hydrogen bonds found in the Watson-Crick base pairs, but two hydrogen

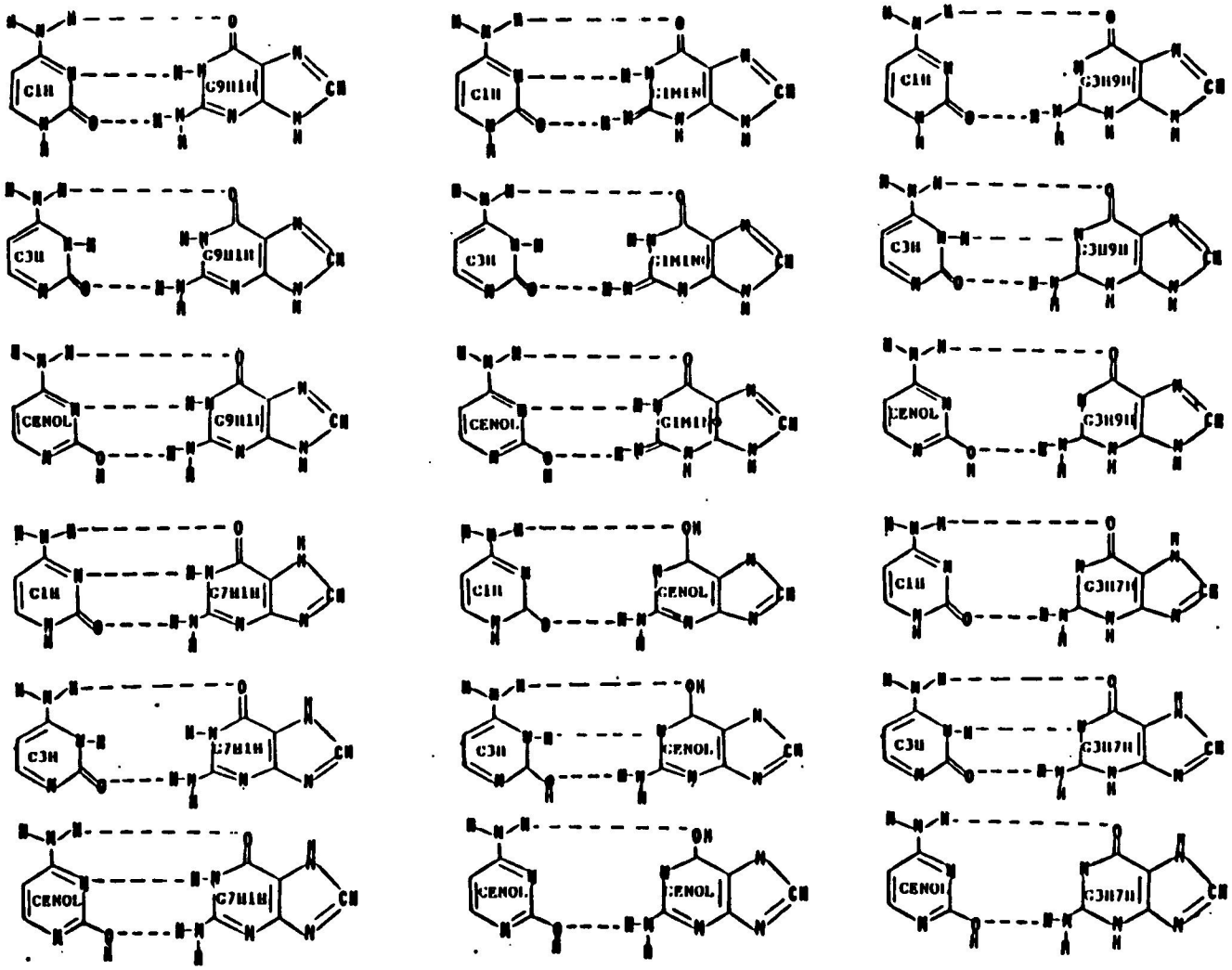


Figure 3-4. Possible tautomeric guanine-cytosine base pairs.

TABLE 3-4. Types of Hydrogen Bonds in the Tautomeric Base Pairs of Guanine and Cytosine

| THREE HYDROGEN BONDS | $E^a$ STO-3G | $\Delta E^b$ | TWO HYDROGEN BONDS | $\Delta E^b$ | $E^a$ STO-3G |
|---|---|---|---|---|---|
| CENOL···G9H1H | -920.0184 | 0 | CENOL···C3H9H | 0 | -919.9593 |
| C1H···G9H1H | -920.0150 | 2.13 | C1H···G3H9H | 3.32 | -919.9540 |
| C1H···G7H1H | -919.9837 | 21.77 | CENOL···G3H7H | 7.46 | -919.9474 |
| CENOL···G7H1H | -919.9826 | 22.46 | CENOL···GENOL | 9.66 | -919.9439 |
| C1H···GIMINO | -919.9643 | 33.94 | C1H···G3H7H | 10.91 | -919.9419 |
| C3H···G3H9H | -919.9604 | 36.39 | C1H···GENOL | 31.37 | -919.9093 |
| CENOL···GIMINO | -919.9568 | 38.65 | C3H···G9H1H | 33.50 | -919.9059 |
| C3H···G3H7H | -919.9480 | 44.17 | C3H···GIMINO | 63.50 | -919.8581 |
| C3H···GENOL | -919.9254 | 58.35 | C3H···G7H1H | 66.26 | -919.8537 |

[a]The total energies (E) in a.u.
[b]The relative energies (ΔE) in kcal/mole
[c]G = Guanine, 9H1H = The hydgogen are located at position 1 and 9 of guanine.
[d]C = Cytosine, 1H = The hydrogen are located at position 1 of cytosine.

bonds were also found in G-C tautomeric base pairs (Figure 3-4) which are usually found in adenine-thymine (A-T) base pairs. The stability of a base pair is closely related to the number of hydrogen bonds present in the base pair. This fact is observed in the melting temperature (Tm) curve of DNA molecule. Due to the number of hydrogen bonds in the G-C base pairs, this base pair is more stable and requires more heat to dissociate than A-T base pair which have only two hydrogen bonds.[8] However, our electronic energy calculations show that the C1H····G9H1H (Watson-Crick) base pair, although it has three hydrogen bonds, it is not the most stable pair. The Cenol···G9H1H base pair is more stable than C1H····G9H1H (Watson-Crick) base pair by only 2.13 kcal/mole (Table 3-4). Elongation is observed upon formation of hydrogen bonds in the majority of the tautomeric base pairs studied (using C1H...G9H1H as a reference). This phenomenon was also observed by Hobza et al.[20] in their base pairing study. The Cenol···G3H9H tautomeric base pair is the most stable pair for the two hydrogen bonded species. It is more stable than the C1H···Genol by 3.32 kcal/mole. Cenol···G3H9H with just two hydrogen bonds is lower in energy by 21.27 kcal/mole than the C3H···Genol pair which has three hydrogen bonds.

Consequently, the number of hydrogen bonds is not the only factor to determine the stability of a G-C base pair.[19] The directional character of these hydrogen bonds may also

play a major role in the hydrogen bonding interaction in nucleic acids. The hydrogen bonding distances for N-H···O ranged from 2.55 to 3.17 A, and the N-H···O ranged from 2.63 to 2.73 A. These hydrogen bonding distances are in the range of values reported by other groups. [1,20,21] We have not taken into account the location and atoms involved in hydrogen bonding to categorize the energies and lengths of specific types of hydrogen bonding interactions because our calculations are at the minimal basis set level.

The energy analysis for the tautomeric base pair of guanine and cytosine is presented in Table 3-5. The C1H···G9H1H base pair represents the most stable interaction when the parent base is C1H, and is a Watson-Crick base pair. The C1H···G9H1H pair is more stable than the C1H···G7H1H tautomeric base pair by 19.54 kcal/mole. The C3H···G3H9H tautomeric base pair is the most stable hydrogen bonding interaction when the parent base is C3H. C3H···G3H9H has three hydroden bonds while C3H···G9H1H has only two hydrogen bonds, and C3H···G3H9H is more stable than the C3H···G9H1H by 34.19 kcal/mole. The Cenol···G9H1H tautomeric base pair is the most stable hydrogen bonding interaction when the parent base is Cenol. It is more stable than the Cenol···G7H1H and Cenol···G3H9H tautomeric base pair by 22.46 and 37.08 kcal/mole, respectively. The stability of Cenol···G9H1H may be attributed to the fact that the separated species Cenol and G9H1H are the most

TABLE 3-5. The Energy Analysis for the Tautomeric Base Pair of Guanine and Cytosine

| | **C1H**[a,b] | $\Delta E^c$ | **C3H**[a,b] | $\Delta E^c$ | **CENOL**[a,b] | $\Delta E^c$ |
|---|---|---|---|---|---|---|
| **G9H1H** | -920.0150 | 0 | -919.9059 | 34.19 | -920.0184 | 0 |
| **G7H1H** | -919.9837 | 19.54 | -919.8537 | 66.95 | -919.9826 | 22.46 |
| **G3H9H** | -919.9540 | 38.27 | -919.9604 | 0 | -919.9593 | 37.08 |
| **G3H7H** | -910.9419 | 45.87 | -919.9480 | 7.78 | -919.9474 | 44.55 |
| **GENOL** | -919.9039 | 69.71 | -919.9254 | 21.77 | -919.9439 | 46.48 |
| **GIMINO** | -919.9643 | 31.81 | -919.8581 | 64.19 | -919.9568 | 38.65 |

[a]The total energies (E) in a.u.
[b]Using STO-3G basis set
[c]The relative energies ($\Delta E$) in kcal/mole
[d]G = Guanine, 9H1H = The hydgogen are located at position 1 and 9 of guanine.
[e]C = Cytosine, 1H = The hydrogen are located at position 1 of cytosine.

stable tautomers of cytosine and guanine at STO-3G level. Therefore, the Cenol···G9H1H base pair has the lowest energy when Cenol is the parent molecule (Table 3-5).

Biological Implication

Heterocyclic molecules in solution frequently yield a mixed population of species in rapid equilibrium. This condition prevails when hydrogen atoms that are attached to nitrogens are able to move to other free nitrogens or keto oxygens within the same molecule.[22,23] The heterocyclic systems of guanine, cytosine, thymine, adenine, and uracil are susceptible to tautomeric changes.[1] Tautomeric forms of the nucleic acids can allow alternative base pairing schemes such as G-U (U=uracil) and A-C which can be important in stabilizing RNA structures[24,25], enzymatic and nonenzymatic codon-anticodon recognition.[12] When tautomerization occurs, it can lead to new hydrogen bonding characteristics within the helix which can lead to biologically active macromolecular conformational changes.

Hydrogen bonding energy analysis for the tautomeric base pairs of guanine and cytosine is shown in Table 3-6. Seven of the eighteen tautomeric base pairs of guanine and cytosine are not energetically favorable relative to the separated species (guanine and cytosine). Therefore, the probability of these seven base pairs participating in biological systems will be low. Genol tautomer participates in four of the eighteen energetically unfavorable tautomeric

TABLE 3-6. Hydrogen Bonding Energy Analysis of the Tautomeric Base Pairs of Guanine and Cytosine

| TAUTOMERIC BASE PAIRS | $\Delta E^{a,b,c}$ | $\Delta E^{a,b,c}/_{HB}{}^{d}$ |
|---|---|---|
| C3H···G3H9H | -31.37 | -10.45 |
| C3H···G3H7H | -30.68 | -10.22 |
| C1H···GIMINO | -17.88 | -5.96 |
| C1H···G3H9H | -9.22 | -4.61 |
| C1H···G3H7H | -8.72 | -4.36 |
| CENOL···C3H9H | -7.15 | -3.57 |
| CENOL···G3H7H | -6.77 | -3.38 |
| CENOL···GIMINO | -7.78 | -2.59 |
| C1H···G9H1H | -5.33 | -1.77 |
| C1H···G7H1H | -4.76 | -1.58 |
| CENOL···G9H1H | -2.07 | -0.690 |
| C3H···GENOL | 0.06 | 0.02 |
| CENOL···G7H1H | 1.31 | 0.44 |
| CENOL···GENOL | 11.98 | 5.99 |
| C1H···GENOL | 28.30 | 14.15 |
| C3H···GIMINO | 30.62 | 15.31 |
| C3H···G9H1H | 44.99 | 22.49 |
| C3H···G7H1H | 58.67 | 29.33 |

[a]The relative energies ($\Delta E$) in kcal/mole.
[b]Using STO-3G basis set
[c]$\Delta E = E_{base\ pair} - (E_{guanine} + E_{cytosine})$
[d]HB = Hydrogen bond
[e]G = Guanine, 9H1H = The hydgogen are located at position 1 and 9 of guanine.
[f]C = Cytosine, 1H = The hydrogen are located at position 1 of cytosine.

base pairs. The Genol is predicted to be the least stable guanine tautomer[11]; therefore, it can have a destabilizing effect on the G-C tautomeric base pairs. Del Bene[26] observed that protonation of guanine stablizes the G-C base pair, but cytosine protonation destabilizes the G-C base pair. We have also observed a destabilizing effect on the base pair when the C3H cytosine tautomer pairs with guanine. This tautomer also participates in four of the eighteen energetically unfavorable tautomeric base pairs. The C3H cytosine tautomer is also the least stable of the cytosine tautomer investigated in this study. Thus, these results suggest that conformational changes which decrease the stability of the helix may occur when the base pairs adopt unusual hydrogen bonding interactions.

$\Delta E^{o}$ (change in internal energy at $0^{o}K$) can be related to $\Delta G$ (change in Gibbs free energy) by the following equation:

$$\Delta G = \Delta H - T\Delta S \qquad (1)$$

$$\Delta G = \Delta E^{o} + \Delta ZPE^{0--->300} + \Delta H^{0-->300} - T\Delta S \qquad (2)$$

$\Delta ZPE^{0--->300}$ and $\Delta H^{0-->300}$ are the changes in the zero point energy and the temperture dependence of $\Delta H^{o}$ from 0 to 300K, respectively.[27] $\Delta E^{o}$ can be divided into three different parts

$$\Delta E^{o} = \Delta E^{o}_{electronic} + \Delta E^{o}_{vibration} + \Delta E^{o}_{rotational}.$$

However, $\Delta E^{o}_{vibration}$ has a small contribution on the overall $\Delta E^{o}$ and should not be considered in the equation and $\Delta E^{o}_{rotational}$ is zero. Tautomeric equilibria can be influenced by pH, substitution at the N1 position, hydrogen bonding, association, base pairing, solvent polarity and other intermolecular interactions within the nucleoside.[15] In biological systems, the solvent plays an important role in determining the predominant species in solution. Therefore, equation (2) should have a solvation factor added.

$$\Delta G = \Delta E^{o} + \Delta ZPE^{0--->300} + \Delta H^{0-->300} - T\Delta S + \text{solvation} \quad (3)$$

Yoshihisa et al.[21] observed that non-Watson-Crick base pairs become preferable in solution because of the stabilization effect of dipole-dipole interaction between the solute and solvent. Therefore, the solvent effect could stabilize these tautomeric base pairs and could change the order of the relative stabilities; the degree of hydration may also affect the helix confomation. If the relative stabilities of the tautomeric base pairs due to the stabilization energies were all to be considered, the C3H···G3H9H could exist and could be expected to mispair during replication.

## CONCLUSIONS

The relative stability of the tautomeric G-C base pairs using STO-3G basis set are:

C3H···G3H9H > C3H···G3H7H > C1H···GIMINO >

C1H···G3H9H* > C1H···G3H7H* > CENOL···G3H9H* >

CENOL···G3H7H* > CENOL···GIMINO > C1H···G9H1H >

C1H···G7H1H > CENOL···G9H1H > C3H···GENOL >

CENOL···G7H1H > CENOL···GENOL* > C1H···GENOL* >

C3H···GIMINO* > C3H···G9H1H* > C3H···G7H1H*

The number of hydrogen bonds between hydrogen bonded base pairs does not determine the overall stability. The order of stability is given with respect to the tautomeric G-C base pairs (two hydrogen bonds*).

To propose a model which can be used in describing the role of the various conformations of DNA, additional calculations need to be performed such as a complete optimization of each tautomeric base pair, use of extented basis sets (6-31G and 6-31G*) to examine the hydrogen bonding patterns of the tautomeric base pairs, stacking, hydration, DNA protein interaction, and helix conformational changes.

## REFERENCES


1. Saenger, W. "Principles of Nucleic Acid Structure", Springer-Verlag:New York, N. Y., 1984; Chapter 6.

2. Sarai, A., and Saito, M.; Int. J. Quantum Chem., **1985**, 28, 399-409.

3. Freese, E.; Proc. Nat. Acad. Sci. USA, **1959**, 45 ,622-633.

4. Barton, J.K.; Chemical and Engineering News, **1988**, 66, 30-42.

5. Watson, J.D. and Crick, F.H.C., Nature, **1953**, 171, 737-738.

6. Drake, J.; Nature, **1969**, 221, 1128-1132.

7. Kornberg, A. "DNA Synthesis", Freeman:San Francisco, 1974.

8. Lehninger, A.L. "Principle of Biochemistry", Worth:New York, 1982.

9. Del Bene, J.E.; J. Phys. Chem., **1983**, 87, 367.

10. C. Singh, and P. Kollman, Quantum Chemistry Program Exchange (QCPE), Program No. 446, Department of Pharmaceutical Chemistry, University of California San Francisco.

11. Jones, P.Y., Bhatia, S.C., and Hall,Jr., J.H., to be submitted to the J. Mol. Struc.

12. (a) Lee, G.C.Y., and Chan, S.I.; J. Am. Chem. Soc., **1972**, 94, 3219-3229.

    (b) Lee, G.C.Y., Prestegard, J.H, and Chan, S.I.; Biochem. Biophys. Res. Commum., 1971, 43, 435.

    (c) Lee, G.C.Y., Prestegard, J.H, and Chan, S.I.; J. Am. Chem. Soc.,**1972**, 94, 951.

13. Miles, H.T.; Proc. Natl. Acad. Sci. U.S.A., **1961**, 47, 791.

14. Wong, P.Y., Wong, K.L., and Kearns, D.R.; Biochem. Biophys. Res. Commun., 1972, 49, 1980.

15. Scanlan, M.J., and Hillier, I.H.; J. Am. Chem. Soc., **1984**, 106, 3737-3745.

16. Schlegel, H., Gund, P., and Fluder, E.; J. Am. Chem. Soc., **1982**, 104, 5347.

17. Czerminski, R., Lesyng, B., and Pohorille, A.; Int. J. Quantum Chem., **1979**, 26, 605-613.

18. Dreyfus, M., Bensaude, G., Dodin, G., and Dubors, J.E.; J. Am. Chem. Soc., **1976**, 98, 9338.

19. (a) Vinogradov, S.N., and Linnell, R.H., "Hydrogen Bonding", Van Nostrand Reinhold: New York, 1971.

    (b) Jeffrey. G.A., and Takagi, S.; Acc. Chem. Res., **1978**, 11, 246-270.

    (c) Kollman, P.A., and Allen, L.C.; Chem. Rev., **1972**, 72, 283-303.

20. Hobza, P., and Sandorfy, C.; J. Am. Chem. Soc., **1987**, 109, 1302-1307.

21. Yoshihisa, O., Hidetsugu, T., Yoshihiro, B., Akihiro, K. and Kichisuke, N.; J. Phys. Chem., **1986**, 90, 4438-4442.

22. Pauling, L. "The Nature of Chemical Bonds", Cornell Univ. Press: Ithaca, N.Y., 1978.

23. Ingold. C.K. "Structure and Mechanism in Organic Chemistry", Cornell Univ. Press: Ithaca, N.Y., 1953.

24. Martin, F.H., Uhlenbeck, O.C., and Doty, P.; J. Mol. Biol., **1971**, 57, 201.

25. Uhlenbeck, O.C., Martin, F.H.,and Doty, P.; J. Mol. Biol., **1971**, 57, 210.

26. Del Bene, J.E.; J. Mol. Struct. (Theochem), **1985**, 124, 201-212.

27. Yanson. I.K., Teplitsky, A.B., and Sukhodub, L.F.; Biopolymers, **1979**, 18, 1149.

# Appendix

# A Study of the Watson-Crick and Non-Watson-Crick Base Pairs

By

Gerald A. Walker,[a] Phelesia Y. Jones,[b] Subhash C. Bhatia,[a] and John H. Hall, Jr.[a,c]

[a]Dolphus E. Milligan Science Research Institute
Atlanta University Center, Inc.
Atlanta, Georgia 30310

[b]Department of Chemistry, Atlanta University
Atlanta, Georgia 30314

[c]School of Geophysical Sciences,
Georgia Institute of Technology
Atlanta, Georgia 30332

## ABSTRACT

*Ab initio* SCF calculations were performed on the substituted and unsubstituted nucleic acid bases, their tautomers, and their hydrogen-bonded base pairs using STO-3G and 4-31G basis sets. Localized molecular orbitals (LMOs) obtained from the Partial Retention of Diatomic Differential Overlap (PRDDO) method are employed to examine the changes occuring between the bonding patterns for the individual bases and the hydrogen-bonded base pairs. Our results suggest that the substituent alters tautomerization in cytosine, and that H-bonding interactions may be altered via the nearest neighbor effect. Our results indicate that these two factors can affect the relative stability of the H-bonded base pairs. A reaction mechanism is proposed to explain the biological behavior and mutagenicty of $N^4$-substituted cytosines.